

## 2 Public-Health-Methoden

---

### 2.1 Epidemiologie

*Oliver Razum, Patrick Brzoska, Matthias Egger*

Die Epidemiologie ist eine Kernwissenschaft für Public Health: Sie ist unentbehrlich, um den Gesundheitszustand auf der Bevölkerungsebene zu beschreiben, Krankheitsursachen und damit Interventionsmöglichkeiten zu identifizieren und deren Wirksamkeit zu messen. Wörtlich übersetzt ist Epidemiologie die Lehre von dem, was „über das Volk kommt“ [von *epi* (gr.): über und *démos* (gr.): Volk]. Sie untersucht die Verteilung von Krankheiten, Todesfällen und anderen gesundheitlichen Ereignissen („Outcomes“) in Bevölkerungen oder Bevölkerungsgruppen, aber auch von Risikofaktoren und schützenden Faktoren (beide werden unter dem Begriff „Expositionen“ zusammengefasst). Die deskriptive Epidemiologie beschreibt dabei die Verteilung von Outcomes und Expositionen, die analytische Epidemiologie schließt aus den Verteilungsmustern auf mögliche Krankheitsursachen und setzt dazu epidemiologische Studiendesigns wie Kohortenstudien und Fall-Kontroll-Studien ein. Bei der Betrachtung der Studienergebnisse stellen EpidemiologInnen systematische Überlegungen zu möglichen Verzerrungen und ihren Folgen sowie zur Ursächlichkeit (Kausalität) der beobachteten Zusammenhänge an. Die Ergebnisse solcher epidemiologischer Studien helfen, präventive Interventionsmaßnahmen zu erarbeiten und diese zu evaluieren.

In diesem Abschnitt betrachten wir zuerst die Rolle der Epidemiologie in Public Health. Anschließend beschäftigen wir uns mit epidemiologischen Verfahren zum Messen und Vergleichen, schauen uns verschiedene epidemiologische Studientypen an und erörtern zum Schluss, wie Schlussfolgerungen aus epidemiologischen Untersuchungen gezogen werden können und welche möglichen Fehlerquellen hier auftreten können.

Schweizerische Lernziele: CPH 5–12

#### 2.1.1 Die Rolle der Epidemiologie in Public Health

##### Epidemiologie – Definition und Überblick

Die Epidemiologie untersucht die Verteilung von gesundheitsrelevanten Ereignissen und Determinanten in Bevölkerungen oder Bevölkerungsgruppen. Solche *gesundheitsrelevanten Ereignisse* – „Outcomes“ – sind v. a. Krankheiten und Todesfälle. Zu den *Determinanten* gehören Risikofaktoren und schützende (*protektive*) Faktoren – EpidemiologInnen sprechen hier allgemein von „Expositionen“. Expositionen können sich aus dem individuellen Verhalten von Menschen ergeben (z. B. Rauchen oder regelmäßiger kör-

perlicher Aktivität), aber auch aus der physikalischen Umwelt (z. B. Zugang zu Gesundheitsdiensten). Mit einer „Bevölkerung“ oder Population im epidemiologischen Sinne können – je nach Situation – alle Menschen eines Landes gemeint sein, aber auch Untergruppen wie z. B. alle Menschen über 65 Jahre oder alle TeilnehmerInnen einer Studie.

Untersuchungsgegenstand der Epidemiologie sind heute nicht nur Infektionskrankheiten, sondern auch nichtübertragbare, chronische Erkrankungen wie etwa der Diabetes mellitus und seine Risikofaktoren. Epidemiologie beschäftigt sich darüber hinaus z. B. auch mit berufsbedingten Erkrankungen und Unfällen. So konnten EpidemiologInnen das gehäufte Auftreten von Blasenkrebs nach einer beruflichen Exposition gegenüber aromatischen Aminen nachweisen (s. Kap. 6.2). Ein weiteres Beispiel ist der Nachweis der Häufung von Verkehrsunfällen unter jungen männlichen Autofahrern jeweils in der Nacht von Freitag und Samstag.

Solche Kenntnisse über die Verteilung von Gesundheitsproblemen und ihren Risikofaktoren in der Bevölkerung ermöglichen es, die jeweilige Größe der Probleme quantitativ zu beschreiben. Hieraus sind dann Rückschlüsse auf die Krankheitsursachen möglich, sodass geeignete Maßnahmen zur Prävention definiert werden können. Schließlich kann auch die Wirksamkeit solcher Maßnahmen evaluiert werden. EpidemiologInnen wenden hierbei deskriptive und analytische Verfahren an.

Die **deskriptive Epidemiologie** beschreibt ein Gesundheitsproblem, indem sie die folgenden Fragen beantwortet:

- Wann treten die Krankheits-/Todesfälle auf? (Verteilung über die Zeit)
- Wo treten die Krankheits-/Todesfälle auf? (geografische Verteilung)
- Wer ist erkrankt? Wie viele Menschen erkranken/versterben? Wer ist exponiert? Wie viele Menschen sind exponiert?

Die Fragen *Wann?*, *Wo?* und *Wer?* (*Time*, *Place*, *Person*) werden als die drei epidemiologischen Fragen bezeichnet. Sie sind die Grundlage des epidemiologischen Arbeitens. Eine Form der deskriptiven Epidemiologie und gleichzeitig Datenquelle für weitere epidemiologische Auswertungen ist die *Gesundheitsberichterstattung* (GBE). Sie umfasst z. B. die Datensätze der Todesursachenstatistik. Diese enthält u. a. Angaben zur Anzahl der Todesfälle, unterschieden (*stratifiziert*) nach Todesursachen, Alter, Geschlecht und Sterbejahr. Details zur GBE finden sich jeweils auf den Websites von *Statistik Schweiz* und der *Gesundheitsberichterstattung des Bundes* (s. Internet-Ressourcen).

Die **analytische Epidemiologie** befasst sich mit der Ermittlung von Risikofaktoren und von Krankheitsursachen. Dazu werden epidemiologische Studiendesigns eingesetzt, bei denen Vergleiche zwischen Populationen angestellt werden. Auch in der analytischen Epidemiologie werden drei Fragen beantwortet (s. a. Kap. 2.1.3 und 2.1.8):

- Besteht eine Assoziation zwischen einem vermuteten Risikofaktor und dem untersuchten Outcome?
- Wie stark ist die Assoziation?
- Ist die beobachtete Assoziation ursächlich (kausal)?

Sind auf diese Weise Risikofaktoren identifiziert, die zum Auftreten des untersuchten Outcomes beitragen (*Attributables Risiko*, s. Kap. 2.1.3), so können geeignete präven-

tive Interventionen entwickelt werden. Deren Wirksamkeit müsste sich durch eine verringerte Häufigkeit des Outcomes zeigen lassen, etwa mit Hilfe von Daten der GBE. In der Realität treten dabei jedoch häufig *Störfaktoren* auf (s. Kap. 2.1.8). Ein wissenschaftlich solider Wirksamkeitsnachweis auf höchstem Evidenzniveau kann nur experimentell durch eine *randomisierte kontrollierte Studie* erbracht werden (s. Kap. 2.1.6).

Zusätzlich zur Unterteilung in deskriptive und analytische Epidemiologie werden innerhalb der Epidemiologie noch verschiedene Themen- und Forschungsbereiche unterschieden. Beispiele hierfür sind die Umweltepidemiologie, die Ernährungsepidemiologie, die Sozialepidemiologie, die klinische Epidemiologie und die molekulare Epidemiologie.

### Einige Meilensteine der Epidemiologie

Das „Denken in Bevölkerungen“ ist keine neue Erfindung. Wichtige Grundprinzipien der Epidemiologie sind schon seit Jahrzehnten oder Jahrhunderten bekannt. Die folgenden Pioniere der Epidemiologie lassen wichtige Ideen und Herangehensweisen erkennen, die in der Epidemiologie eine große Rolle spielen:

**John Graunt** (1620–1674), ein englischer Kaufmann, analysierte die Listen aller Todesfälle, die schon damals in London geführt wurden – ähnlich, wie dies in der *Todesursachenstatistik* heute noch geschieht. Graunt stellte fest, dass Kinder ein höheres Sterberisiko als Erwachsene hatten, dass das Sterberisiko bei Männern höher war als bei Frauen, und dass die Sterblichkeit in London höher lag als auf dem Lande. Hieraus schlussfolgerte er, dass die Risiken für Krankheit und Tod nicht zufällig und nicht gleichmäßig in der Bevölkerung verteilt sind. Diese Erkenntnis mag banal erscheinen, sie ist aber Grundlage jeglicher Epidemiologie. Würden Krankheiten und Todesfälle rein zufällig auftreten, so könnte man keine Risikofaktoren identifizieren (wie z. B. das Rauchen als Risikofaktor für Lungenkrebs) oder Bevölkerungsgruppen benennen, die ein erhöhtes Risiko aufweisen (wie etwa allein stehende, ältere Männer für Suizid).

Der englische Arzt **John Snow** (1813–1858) untersuchte die großen *Cholera-Ausbrüche*, die im 19. Jahrhundert in den Städten zu Tausenden von Toten führten. Damals waren Erreger und Übertragungsweg der Seuche noch unbekannt. Snow zeigte mit deskriptiven und analytischen epidemiologischen Methoden, dass kontaminiertes Trinkwasser eine wesentliche Rolle bei der Übertragung der Cholera in London spielte. Viele Jahre vor der Kultivierung des Erregers *Vibrio cholerae* durch *Robert Koch* konnte er aus seinen Studienergebnissen wirksame Präventionsmaßnahmen ableiten.

Der deutsche Mediziner und Begründer der Zellpathologie **Rudolf Virchow** (1821–1902) leistete Pionierarbeit auf dem Gebiet der Sozialepidemiologie. Virchow beobachtete während einer *Hungertyphus-Epidemie* (Typhus exanthematicus; Läusefleckfieber) in Oberschlesien, dass Armut krank macht und Krankheit somit auch gesellschaftliche Ursachen hat. Medikamente allein reichen nicht, um den Gesundheitszustand der Bevölkerung zu verbessern, solange es an bezahlter Arbeit, Bildung und sozialer Absicherung mangelt. Virchow prägte 1848 den Satz „Die Medizin ist eine soziale Wissenschaft, und die Politik ist weiter nichts als Medizin im Großen“.

Der Epidemiologe **Richard Doll** (1912–2005) und der Statistiker **Austin Bradford Hill** (1897–1991) führten die *British Doctors Study* durch, eine modellhafte Kohortenstudie

(s. Kap. 2.1.5) zum Einfluss des Rauchens auf die Sterblichkeit an Lungenkrebs und anderen Erkrankungen. Im Jahr 1951 rekrutierten Doll und Hill mehr als 34.000 Ärzte aus dem britischen Ärztereister und fragten nach deren Rauchgewohnheiten. Sie beobachteten die Ärzte über viele Jahre und verglichen unter anderem die Häufigkeit von Todesfällen an Lungenkrebs und Herz-Kreislauf-Krankheiten unter exponierten und nicht exponierten Ärzten. Beide trugen dazu bei, dass *Rauchen als Risikofaktor für Lungenkrebs* erkannt wurde, lange bevor die Mechanismen der Krebsentstehung auf zellulärer Ebene verstanden wurden. Hill war darüber hinaus einer der Pioniere auf dem Gebiet der *randomisierten Studien* (s. Kap. 2.1.6) und entwickelte die nach ihm benannten *Bradford-Hill-Kriterien für Kausalität* (s. Kap. 2.1.7).

## 2.1.2 Epidemiologische Verfahren zum Messen und Vergleichen

### Häufigkeitsmaße für Expositionen und Outcomes

Zur Untersuchung der Häufigkeit von *Expositionen* und *Outcomes* nutzt die Epidemiologie verschiedene deskriptive Maßzahlen.

**Absolute Zahl:** Die grundlegendste deskriptive Maßzahl ist die absolute Zahl. Sie gibt die Anzahl von Person an, die einer bestimmten Exposition ausgesetzt sind oder einen bestimmten Outcome aufweisen. Die absolute Zahl ist eine wichtige Grundlage der Gesundheitsberichterstattung und wird aus unterschiedlichen routinemäßig erhobenen Daten, amtlichen Statistiken oder epidemiologischen *Surveys* gewonnen. So zeigt z.B. die amtliche Pflegestatistik, dass am 15. Dezember 2009 die absolute Zahl an Menschen, die nach der Definition des *Sozialgesetzbuchs XI* in Nordrhein-Westfalen pflegebedürftig waren, 509.145 betrug. In Hessen waren zum gleichen Zeitpunkt 186.893 Menschen pflegebedürftig.

**Prävalenz:** Ein Nachteil absoluter Zahlen ist, dass sie keinen Vergleich zwischen einzelnen Regionen oder verschiedenen Zeitpunkten erlauben. So kommt die höhere Zahl der Pflegebedürftigen im deutschen Bundesland Nordrhein-Westfalen vermutlich dadurch zustande, dass die Bevölkerung dort größer ist als im Bundesland Hessen. Ein Vergleich der absoluten Zahlen (*Zähler*) wird daher erst dann möglich, wenn sie in Bezug zur Bevölkerungsgröße (*Nenner*) der jeweiligen Regionen gesetzt werden. (Auf mögliche Altersstruktureffekte wollen wir an dieser Stelle nicht eingehen [s. hierzu Kap. 2.2.2].)

Die daraus resultierende Maßzahl heißt **Punktprävalenz**. Sie beschreibt den Anteil der Exponierten bzw. den Anteil derjenigen, die einen bestimmten Outcome aufweisen (zum Beispiel pflegebedürftig sind), jeweils zu einem definierten Zeitpunkt. Dieser Anteil wird oft pro 100 Personen, manchmal auch pro 1.000, 10.000 oder 100.000 Personen der Gesamtbevölkerung angegeben:

$$\text{Punktprävalenz} = \frac{\text{Personen m. Exposition bzw. Outcome zu einem definierten Zeitpunkt}}{\text{Gesamtbevölkerung zum gleichen Zeitpunkt}} (\cdot 100)$$

Die Punktprävalenz von Pflegebedürftigkeit am 15. Dezember 2009 betrug damit in Nordrhein-Westfalen, wo zu diesem Zeitpunkt 17.872.763 Menschen lebten,

$$509.145 / 17.872.763 \cdot 100 = 2,8 \text{ pro } 100 \text{ Personen.}$$

Insgesamt waren also 2,8% der Menschen in Nordrhein-Westfalen pflegebedürftig. In Hessen lag die Punktprävalenz zum gleichen Zeitpunkt bei 3,1%:

$$186.893 / 6.061.951 \cdot 100 = 3,1 \text{ pro } 100 \text{ Personen}$$

Da sich die Punktprävalenz lediglich auf einen einzigen Zeitpunkt bezieht, stellt sie eine Momentaufnahme dar. Sie ist daher nicht als Maßzahl für Erkrankungen mit einer kurzen Dauer (z. B. Durchfallerkrankungen oder Erkältungen) geeignet. In solchen Fällen wird als alternatives Prävalenzmaß die **Periodenprävalenz** berechnet, die sich auf einen Zeitraum (etwa einen Monat oder ein Jahr) bezieht. Im Nenner der Periodenprävalenz befinden sich alle Fälle zu Beginn des betrachteten Zeitraums sowie alle in diesem Zeitraum neu aufgetretenen Fälle:

$$\text{Periodenprävalenz} = \frac{\text{Erkrankte zu Beginn eines Zeitraums} + \text{Neuerkrankte im Zeitraum}}{\text{Mittlere Bevölkerung im Zeitraum}} (\cdot 100)$$

Als mittlere Bevölkerung im Zeitraum wird in der Regel der Durchschnitt aus der Bevölkerungszahl zu Beginn und zum Ende des Betrachtungszeitraums angegeben.

Einige epidemiologische Lehrbücher verwenden bei der Formel der Punkt- und Periodenprävalenz im Nenner nur die Bevölkerung „unter Risiko“ (*at risk*). Die Bezugsbevölkerung besteht in diesem Fall nur aus denjenigen, die den Outcome ausbilden können. Eine solche Definition ist z. B. bei Fragestellungen sinnvoll, die sich mit geschlechtsspezifischen Krankheiten wie Prostatakrebs oder Gebärmutterhalskrebs beschäftigen.

**Inzidenz:** Die Periodenprävalenz ist eine statische Maßzahl. Die Inzidenz erlaubt es hingegen, Veränderungen innerhalb eines Zeitraums abzubilden. Das geschieht, indem der Zähler nur die neu aufgetretenen (*inzidenten*) Fälle eines bestimmten Zeitraums berücksichtigt. In der Regel wird die Inzidenz nur für Outcomes (z. B. Erkrankungen), seltener für Expositionen verwendet. In der Epidemiologie lassen sich verschiedene Inzidenzmaße unterscheiden.

Die **kumulative Inzidenz** (auch als *Inzidenzrisiko* bezeichnet) ist die Wahrscheinlichkeit, mit der eine Person in einem bestimmten Zeitraum erkrankt. Sie ist das Verhältnis der Zahl an Neuerkrankungen in einem definierten Zeitraum zur Zahl der Bevölkerung unter Risiko zu Beginn des Zeitraums und wird meist pro 1.000 oder 100.000 Personen angegeben.

$$\text{kumulative Inzidenz} = \frac{\text{Neuerkrankte in einem definierten Zeitraum}}{\text{Bevölkerung unter Risiko zu Beginn des Zeitraums}}$$

Die kumulative Inzidenz, ist eine geeignete Maßzahl, wenn Veränderungen innerhalb der Bevölkerung unter Risiko (durch Zu- und Abwanderungen sowie durch Geburten und Sterbefälle) vernachlässigt werden können. Da dies bei vielen epidemiologischen Fragestellungen jedoch nicht der Fall ist und die Bevölkerung unter Risiko im definierten Zeitraum genauer abgebildet werden muss, wird statt der kumulativen Inzidenz die **Inzidenzrate** berechnet. Auch hierbei werden Neuerkrankte in einem definierten Zeitraum betrachtet. Im Nenner befindet sich dann aber die mittlere Bevölkerung unter Risiko:

$$\text{Inzidenzrate} = \frac{\text{Neuerkrankte in einem definierten Zeitraum}}{\text{Mittlere Bevölkerung unter Risiko im gleichen Zeitraum}}$$

Die *mittlere Bevölkerung unter Risiko* ist meist der Durchschnitt aus der Bevölkerung unter Risiko zu Beginn und zum Ende des Betrachtungszeitraums. Wie die kumulative Inzidenz wird die Inzidenzrate oft pro 1.000 oder 100.000 Personen angegeben. Im Jahr 2006 wurden in Deutschland z. B. 7.360 Hautkrebs-Neuerkrankungen bei Männern gemeldet. Die mittlere männliche Bevölkerung umfasste in diesem Jahr 40.317.807 Personen. Hieraus lässt sich eine Inzidenzrate von  $(7.360/40.317.807) \cdot 100.000 = 18,3$  pro 100.000 Männern errechnen. Dies bedeutet, dass im Jahr 2006 in Deutschland pro 100.000 männliche Personen etwa 18 Männer neu an Hautkrebs erkrankten.

In epidemiologischen Studien setzt ein solches Vorgehen voraus, dass alle Studienteilnehmer zeitgleich in die Studie aufgenommen werden. In der Praxis ist das oft nicht möglich, da die Rekrutierung meist einen längeren Zeitraum beansprucht. Außerdem nehmen nicht alle Personen, die zu Beginn in eine Studie eingeschlossen werden, auch bis zum Ende daran teil. Manche von ihnen entwickeln den Outcome (d. h. sie erkranken), andere Teilnehmer wollen nicht länger an der Studie teilnehmen, oder man weiß nichts mehr über ihren Verbleib. Sie alle scheiden aus der Studie aus und zählen nicht länger zur Bevölkerung unter Risiko. Um die unterschiedlichen Zeiträume zu berücksichtigen, die Personen zur Bevölkerung unter Risiko gehören, wird in epidemiologischen Studien bei der Berechnung der Inzidenzrate im Nenner statt der mittleren Bevölkerung häufig die *Personenzeit unter Risiko* verwendet. In diesem Fall heißt die Inzidenzrate auch **Inzidenzdichte**:

$$\text{Inzidenzdichte} = \frac{\text{Neuerkrankte in einem definierten Zeitraum}}{\text{Personenzeit unter Risiko im gleichen Zeitraum}}$$

Sie wird meist pro 1.000 oder 100.000 Personenjahre angegeben.

Spezielle Inzidenzmaße für die Untersuchung der Sterblichkeit sind die Mortalität und die Letalität (s. a. Kap. 2.2.3). Die **Mortalität** (auch Mortalitätsrate oder Sterbeziffer) bezeichnet die Zahl der Gestorbenen in einem bestimmten Zeitraum (*inzidente Sterbefälle*) im Verhältnis zur mittleren Bevölkerung in dieser Zeit. Die Mortalität ist eine Maßzahl, die häufig in der Gesundheitsberichterstattung verwendet und meistens pro 100.000 Personen der Bevölkerung angegeben wird:

$$\text{Mortalitätsrate} = \frac{\text{Gestorbene innerhalb eines Zeitraums}}{\text{Mittlere Bevölkerung unter Risiko im gleichen Zeitraum}} (\cdot 100.000)$$

Im Jahr 2009 sind in Deutschland z. B. 216.128 Menschen an bösartigen Tumoren gestorben. Die mittlere Bevölkerung betrug im gleichen Jahr 81.874.770 Personen. Daraus lässt sich eine Mortalitätsrate von  $(216.128/81.874.770) \cdot 100.000 = 263,97$  pro 100.000 Personen errechnen. Im Jahr 2009 starben in Deutschland also etwa 264 von 100.000 Menschen an bösartigen Tumoren.

Die **Letalität** wird meist in Prozent ausgedrückt und bezeichnet die Zahl der in einem definierten Zeitraum an einer Krankheit Gestorbenen im Verhältnis zur Zahl der Erkrankten im gleichen Zeitraum:

$$\text{Letalität} = \frac{\text{An einer Krankheit Gestorbene innerhalb eines Zeitraums}}{\text{Alle von der Krankheit betroffenen Personen im Zeitraum}} (\cdot 100)$$

Beispiel: Im Jahr 1976 erkrankten 182 Veteranen der *American Legion*, die sich in Philadelphia zu einem Treffen versammelt hatten, an einer bis dahin unbekannten Form der Lungenentzündung. Neunundzwanzig der 182 erkrankten Personen verstarben (Letalität 16%). Der Erreger erhielt den Namen *Legionella pneumophila*, und die Krankheit wurde fortan als Legionärskrankheit bekannt.

### 2.1.3 Assoziationsmaße für Expositionen und Outcomes

Aufgabe der Epidemiologie ist es nicht nur, die Häufigkeit von Expositionen und Outcomes in einer Bevölkerung zu untersuchen, sondern auch die Stärke des Zusammenhangs (*Assoziation*) zwischen ihnen zu bestimmen. Dafür stehen unterschiedliche **Assoziationsmaße** zur Verfügung. Um einen solchen Zusammenhang zu berechnen, nutzt die Epidemiologie häufig die sogenannte *Vier-Felder-Tafel* als Hilfsmittel. Ein hier festgestellter Zusammenhang sagt jedoch noch nichts über eine mögliche Ursache-Wirkungs-Beziehung aus (s. Kap. 2.1.8).

**Vier-Felder-Tafel:** Eine Vier-Felder-Tafel (*Kontingenztafel*) stellt die Häufigkeit einer Exposition in Abhängigkeit zu einem Outcome dar (s. Tab. 2.1).

**Tab. 2.1:** Gerüst einer Vier-Felder-Tafel.

		Outcome		Summe
		Ja	Nein	
Exposition	Ja	a	b	a+b
	Nein	c	d	c+d
Summe		a+c	b+d	a+b+c+d

Die vier Felder bezeichnen hierbei:

- die exponierten Personen, bei denen der Outcome aufgetreten ist (*a*),
- die exponierten Personen, bei denen der Outcome nicht aufgetreten ist (*b*),
- die nicht exponierten Personen, bei denen der Outcome aufgetreten ist (*c*),
- die nicht exponierten Personen, bei denen der Outcome nicht aufgetreten ist (*d*).

Weitere wichtige Informationen der Vier-Felder-Tafel ergeben sich aus den fünf Randsummen. Sie geben Auskunft über

- alle exponierten Personen (*a+b*),
- alle nicht exponierten Personen (*c+d*),
- alle Personen, bei denen der Outcome aufgetreten ist (*a+c*),
- alle Personen, bei denen der Outcome nicht aufgetreten ist (*b+d*),
- alle Personen, die untersucht wurden (*a+b+c+d*).



**Relatives Risiko und Relative Rate:** Aus der Vier-Felder-Tafel können die *kumulativen Inzidenzen* unter den exponierten Personen ( $a/[a+b]$ ) und den nicht exponierten Personen ( $c/[c+d]$ ) abgelesen werden. Das Verhältnis der beiden Inzidenzen ist ein Maß für die Stärke des Zusammenhangs zwischen Exposition und Outcome. Es wird als **Relatives Risiko (RR)** bezeichnet:

$$\text{Relatives Risiko} = \frac{\text{Kumulative Inzidenz unter den Exponierten}}{\text{Kumulative Inzidenz unter den nicht Exponierten}} = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Ist das Relative Risiko größer als 1, weist das auf ein höheres Risiko der Exponierten hin, den Outcome auszubilden (z. B. zu erkranken). Bei einem Relativen Risiko kleiner als 1 ist das Risiko für die Exponierten, den Outcome auszubilden, geringer als unter den nicht Exponierten. Je weiter das Relative Risiko von 1 entfernt ist, desto größer ist der Unterschied zwischen Exponierten und nicht Exponierten und damit auch der Zusammenhang zwischen Exposition und Outcome.

Liegen statt kumulativer Inzidenzen *Inzidenzraten* vor, ist es strenggenommen nicht korrekt, von einem Relativen Risiko zu sprechen, da man Zähler und Nenner nicht als Wahrscheinlichkeiten interpretieren kann. Der Quotient aus der Inzidenzrate unter den Exponierten und der Inzidenzrate unter den nicht Exponierten wird stattdessen als **Relative Rate** bezeichnet. Er wird wie das Relative Risiko berechnet:

$$\text{Relative Rate} = \frac{\text{Inzidenzrate unter den Exponierten}}{\text{Inzidenzrate unter den nicht Exponierten}}$$

In der *British Doctors Study* betrug z. B. nach 20 Jahren Beobachtungszeit die Inzidenzdichte für Speiseröhrenkrebs 14 pro 100.000 bei Zigarettenrauchern und 3 pro 100.000 bei Nichtrauchern. Die relative Rate berechnet sich folgendermaßen:

$$\text{Relative Rate} = (14/100.000) / (3/100.000) = 4,7$$

Raucher haben demnach eine fast fünfmal so hohe Rate wie Nichtraucher, an Speiseröhrenkrebs zu versterben.

Um Relative Risiken oder Relative Raten berechnen zu können, muss die kumulative Inzidenz bzw. die Inzidenzrate unter den Exponierten und nicht Exponierten bekannt sein. In Kohorten- und randomisierten kontrollierten Studien (s. Kap. 2.1.5 und Kap. 2.1.6) ist dies der Fall. Bei Studientypen, wo dies nicht möglich ist (z. B. Querschnitt- oder Fall-Kontroll-Studien), wird stattdessen als alternatives Assoziationsmaß die Odds Ratio verwendet.

**Odds Ratio:** Die **Odds Ratio (OR)** basiert – anders als das Relative Risiko – nicht auf der kumulativen Inzidenz, sondern auf der Chance (im Englischen als Odds bezeichnet) von Exponierten und nicht Exponierten, dass ein Outcome auftritt. Die Chance ist hierbei ein Quotient, bestehend aus der Wahrscheinlichkeit für einen Outcome und seiner Gegenwahrscheinlichkeit:

$$\text{Odds} = \frac{\text{Wahrscheinlichkeit für einen Outcome}}{1 - (\text{Wahrscheinlichkeit für einen Outcome})}$$



Die Odds Ratio setzt also die Chancen von Exponierten und nicht Exponierten zueinander ins Verhältnis. Sie ist damit ein Quotient zweier Quotienten und wird deshalb manchmal auch als *Quotenquotient* oder *Chancenverhältnis* bezeichnet. Die beiden Chancen lassen sich leicht aus einer Vier-Felder-Tafel ablesen ( $a/b$  und  $c/d$ ). Durch die Kehrwertregel kann man die Formel für die Odds Ratio schließlich wie hier dargestellt vereinfachen:

$$\text{Odds Ratio} = \frac{\text{Odds unter den Exponierten}}{\text{Odds unter den nicht Exponierten}} = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \cdot d}{b \cdot c}$$

Interpretieren lässt sich die Odds Ratio ähnlich wie das Relative Risiko oder die Relative Rate. Werte über 1 weisen auf eine höhere Chance, Werte unter 1 auf eine geringere Chance der Exponierten hin, den Outcome auszubilden. Je weiter eine Odds Ratio nach unten oder oben von der 1 abweicht, desto stärker ist der Zusammenhang zwischen Exposition und Outcome. Ist der Outcome selten, liegen die Odds Ratio oder das Relative Risiko eng beieinander. Wenn der Outcome hingegen häufig ist, können Odds Ratio und Relatives Risiko stark voneinander abweichen.

Die Vier-Felder-Tafel in Tab. 2.2 illustriert die Berechnung und Interpretation der Odds Ratio am Beispiel einer Fall-Kontroll-Studie zum Einfluss von Neuroleptika (*Exposition*) auf die Entstehung einer venösen Thromboembolie (VTE; *Outcome*).

**Tab. 2.2:** Vier-Felder-Tafeln zum Zusammenhang von Neuroleptikaeinnahme und venöser Thromboembolie (VTE).

		Outcome (VTE)			Summe	
		Ja		Nein		
Exposition (Neuroleptikaeinnahme)	Ja	2.126	a	b	4.752	a+b 6.878
	Nein	23.406	c	d	84.739	c+d 108.145
Summe		25.532	a+c	b+d	89.491	a+b+c+d 115.023

**Odds Ratio** =  $(2.126 \cdot 84.739) / (4.752 \cdot 23.406) = 1,62$

Quelle der Originaldaten: Parker C, Coupland C, Hippisley-Cox J. Antipsychotic drugs and risk of venous thromboembolism: nested case-control study. *British Medical Journal* 2010; 341: c4245.

Die Autoren untersuchten insgesamt 115.023 Personen, die in Hausarztpraxen in Großbritannien registriert waren. Hiervon entwickelten 25.532 PatientInnen zwischen 1996 und 2007 eine VTE (*Fälle*), bei den übrigen 89.491 Personen kam es nicht zu einer VTE (*Kontrollen*). Insgesamt nahmen 6.878 Personen Neuroleptika ein (*Exponierte*), 108.145 Personen taten dies nicht (*nicht Exponierte*). Von den Neuroleptika-NutzerInnen litten 2.126 an einer VTE, 4.752 nicht. Die Odds einer VTE betrug daher unter den Exponierten  $2.126 / 4.752$ . Von denjenigen, die keine Neuroleptika einnahmen, litten 23.406 an einer VTE. Ihre Odds lag damit bei  $23.406 / 84.739$ . Um hieraus die Odds Ratio zu errechnen, wird ein Quotient aus beiden Odds gebildet:

$$\text{OR} = (2.126 / 4.752) / (23.406 / 84.739) = (2.126 \cdot 84.739) / (4.752 \cdot 23.406) = 1,6$$

Demnach ist also die Chance bei Personen, die Neuroleptika einnehmen, eine VTE zu bekommen, 1,6-mal so hoch wie bei Personen, die keine Neuroleptika einnehmen.

**Attributables Risiko:** Besteht eine Ursache-Wirkungs-Beziehung zwischen Exposition und Outcome (s. Kap. 2.1.8), erlaubt das **attributable Risiko (AR)** den Anteil bei den neuen Erkrankungen zu ermitteln, der auf die Exposition zurückzuführen ist. Das attributable Risiko wird meist in Prozent angegeben und wie folgt berechnet:

$$\text{Attributables Risiko} = \frac{\left( \text{Kum. Inzidenz unter den Exponierten} \right) - \left( \text{Kum. Inzidenz unter den nicht Exponierten} \right)}{\text{Kum. Inzidenz unter den Exponierten}} (\cdot 100)$$

In der oben erwähnten Analyse der *British Doctors Study* lässt sich das attributable Risiko zum Zusammenhang von Rauchen und Speiseröhrenkrebs-Mortalität folgendermaßen berechnen:

$$(14 / 100.000 - 3 / 100.000) / (14 / 100.000) \cdot 100 = 78,6\%$$

Dies bedeutet, dass 78,6% aller Sterbefälle durch Speiseröhrenkrebs auf Tabakkonsum zurückzuführen sind.

### 2.1.4 Validität und Reliabilität

Die Güte eines Messverfahrens wird durch seine Validität und Reliabilität bestimmt. Die **Validität** (Gültigkeit) bezeichnet das Ausmaß, in dem ein Messverfahren das misst, was es messen soll. Eine Messung gilt dann als valide, wenn ihr Ergebnis der Realität entspricht. Der Begriff der **Reliabilität** (Wiederholbarkeit) beschreibt das Ausmaß, in dem Messwerte replizierbar sind. Ein Messverfahren oder eine Messung gelten also dann als reliabel, wenn zu unterschiedlichen Messzeitpunkten oder bei Messwiederholungen gleiche Messwerte ermittelt werden. Die Web-Abb. 2.1.1 auf unserer Lehrbuch-Homepage zeigt schematisch am Beispiel der Bestimmung des Körpergewichts die Auswirkungen hoher bzw. geringer Validität und Reliabilität eines Messverfahrens auf das Verhältnis zwischen den gemessenen Werten und dem wahren (aber unbekannten) Wert.

In der Epidemiologie – etwa bei der Untersuchung von Ursache-Wirkungs-Beziehungen – wird darüber hinaus auch noch zwischen interner und externer Validität unterschieden. Diese Begriffe beziehen sich jedoch nicht auf die oben beschriebene Eigenschaft von Messungen und Messverfahren, sondern auf die Qualität der Studie und die Anwendbarkeit der Resultate. Eine hohe **interne Validität** liegt dann vor, wenn in einer Studie die beobachtete Ausprägung eines Outcomes allein auf die Exposition zurückzuführen ist und Alternativerklärungen für das Vorliegen oder die Höhe der gefundenen Effekte weitestgehend ausgeschlossen werden können. Dabei sinkt die interne Validität mit der steigenden Anzahl von plausiblen alternativen Erklärungen aufgrund von Fehlern und Verzerrungen. Bevor eine Untersuchung als intern valide bezeichnet werden kann, müssen deshalb Störgrößen als Ursache für einen beobachteten Zusammenhang ausgeschlossen werden (s. Kap. 2.1.8). Der Begriff der **externen Validität** bezeichnet die

Anwendbarkeit oder Generalisierbarkeit der Studienergebnisse über die StudienteilnehmerInnen hinaus auf andere Populationen. Eine geringe externe Validität findet man bei unnatürlichen Untersuchungsbedingungen und bei geringer Repräsentativität der untersuchten Stichprobe. Eine hohe interne Validität ist Voraussetzung für eine hohe externe Validität.

### 2.1.5 Epidemiologische Studientypen

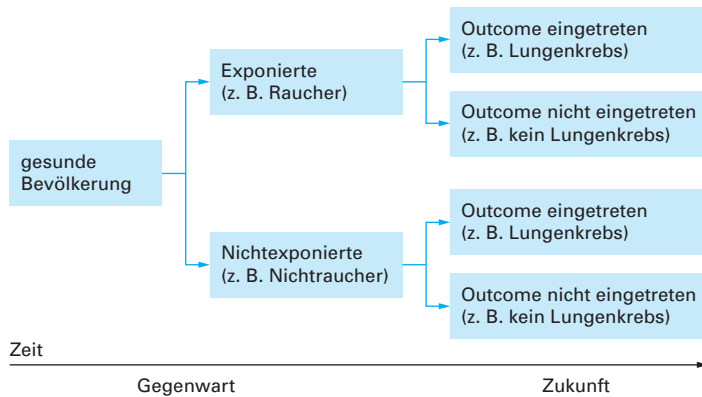
In der analytischen Epidemiologie werden Bevölkerungen oder Bevölkerungsgruppen im Hinblick auf die interessierenden Expositionen und Outcomes verglichen, um die oben erwähnten drei Fragen nach dem Vorhandensein, der Stärke und der Kausalität einer Assoziation (s. Kap. 2.1.1) zu beantworten. Analytisch-epidemiologische Studien im engeren Sinne sind beobachtende Studien. Hier wird die Exposition nicht von den Forschenden zugeteilt, sondern durch sie beobachtet. Die gängigsten Designs sind *Querschnittstudie*, *Kohortenstudie* und *Fall-Kontroll-Studie*. Experimentelle Studien, bei denen ForscherInnen die Exposition zuteilen, werden in Kap. 2.1.6 besprochen.

#### Querschnittstudien

In einer Querschnittstudie werden alle Variablen, deren Assoziation untersucht werden soll, zum gleichen Zeitpunkt erhoben. Das erlaubt eine schnelle Durchführung, kann aber zu schwer interpretierbaren Ergebnissen führen. Da eine zeitliche Achse fehlt, bleibt oft unklar, welche Variable die Exposition ist und welche der Outcome. Wir stellen uns hierzu als Beispiel eine Studie vor, bei der in einem Krankenhaus bei allen PatientInnen der Cholesterinspiegel gemessen wird. Gleichzeitig wird ermittelt, ob die/der PatientIn eine Krebserkrankung hat oder nicht. Es findet sich nun ein statistischer Zusammenhang zwischen einem niedrigen Cholesterinspiegel und einer Krebserkrankung. Nun wäre es falsch, auf der Basis solcher Querschnittdaten den Schluss zu ziehen, dass ein niedriger Cholesterinspiegel ein Risikofaktor für Krebserkrankungen ist. Da die Studie keine zeitlichen Informationen erhoben hat, könnte es ebenso gut sein, dass eine Krebserkrankung zu einem niedrigen Cholesterinspiegel führt – etwa, weil die Betroffenen den Appetit verlieren und nicht mehr genügend essen. Tatsächlich erscheint das sogar als die plausiblere Interpretation dieser Ergebnisse. Mit dem gewählten Querschnittsdesign ist eine Klärung jedoch nicht möglich. Hierzu müsste eine Kohortenstudie durchgeführt werden. Untersuchungen, bei denen lediglich die Prävalenz *eines* Risikofaktors oder *einer* Erkrankung zu einem bestimmten Zeitpunkt gemessen wird, werden ebenfalls als Querschnittstudien oder auch als „Prävalenzstudien“ bezeichnet. In diesem Fall handelt es sich jedoch um deskriptive – und nicht um analytische – Studien.

#### Kohortenstudien

Kohortenstudien werden auch als prospektive (d.h. in die Zukunft schauende) oder longitudinale (sich über einen Zeitraum erstreckende) Studien bezeichnet. Sie beginnen mit einer Gruppe von *Exponierten* und einer Gruppe von *Nichtexponierten*, die alle vom zu untersuchenden Outcome frei – also im Hinblick auf diesen Outcome gesund – sein müssen. Beide Gruppen werden über einen bestimmten Zeitraum nachverfolgt.



**Abb. 2.1:** Schema einer Kohortenstudie.

In jeder der beiden Gruppen wird die Inzidenz des Outcomes berechnet (s. Abb. 2.1). Das ist in Kohortenstudien möglich, da sowohl die Zahl der Fälle als auch die Zahl der Personen unter Risiko bzw. deren Personenzzeit (s. Kap. 2.1.2) bekannt sind. Als Assoziationsmaß dient das *Relative Risiko* (RR) oder die *Relative Rate* (s. Kap. 2.1.3).

Kohortenstudien sind besonders geeignet, wenn

- mehrere Outcomes einer Exposition untersucht werden sollen (z.B. das Risiko von RaucherInnen im Vergleich zu NichtraucherInnen, verschiedene Krebsarten zu entwickeln)
- der Outcome häufig ist
- die Exposition selten ist
- der Expositionsstatus sich über die Zeit verändert (dies lässt sich durch Angabe der exponierten bzw. nicht exponierten Personenzzeit berücksichtigen)

Viele chronische, nichtübertragbare Erkrankungen haben eine lange *Latenzzeit*, d.h. es existiert ein Zeitraum zwischen der Exposition und dem Auftreten des Outcomes. Daher dauern Kohortenstudien, die Ursachen solcher Krankheiten untersuchen, oft Jahre oder Jahrzehnte. Wird dagegen z.B. das Risiko der Entwicklung einer Fehlbildung als Folge einer Exposition während der Schwangerschaft erforscht, so muss die dazu durchgeführte Kohortenstudie nur auf die Dauer der Schwangerschaft angelegt sein.

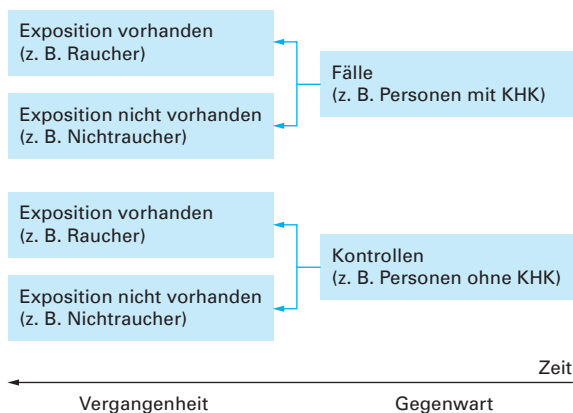
**Historische Kohortenstudien:** Eine Kohortenstudie kann auch so angelegt werden, dass sie sich von einem Zeitpunkt in der Vergangenheit bis in die Gegenwart erstreckt. Das ist möglich, wenn für eine Personengruppe entsprechende Expositions- und Gesundheitsdaten aus der Vergangenheit vorliegen. Solche „historischen Kohortenstudien“ werden z.B. in der Arbeitsepidemiologie durchgeführt, wenn in einem Unternehmen Unterlagen über Expositionen am Arbeitsplatz (z.B. zum Arbeiten mit einem Lösungsmittel) für die gesamte Beschäftigungsdauer vorliegen. Außerdem muss für alle Beschäftigten ermittelt werden, ob der Outcome (etwa eine Krebserkrankung) eingetreten ist. Die nicht exponierte Vergleichsgruppe sollte aus dem gleichen oder einem vergleichbaren, anderen Betrieb kommen, damit der sozioökonomische Status ähnlich ist. Eine

Stichprobe aus der Allgemeinbevölkerung als Vergleichsgruppe kann zu einer Verzerrung führen, da ihr Gesundheitszustand im Schnitt schlechter sein kann als der einer arbeitenden Population (*Healthy-worker-Effekt*). Ein Beispiel für eine historische Kohortenstudie ist die *Swiss National Cohort*. Hierbei wurden die Daten, die Ende 1990 in einer Volkszählung (u. a. zum Bildungsstand der Schweizer Bevölkerung) erhoben wurden, mit der Statistik der Todesfälle der Jahre 1991 bis 2008 verlinkt.

**Die „SAPALDIA Kohorte“ und die „Nationale Kohorte“:** Im Normalfall wird vor dem Beginn einer Kohortenstudie entsprechend der Forschungsfrage genau festgelegt, welche Expositionen und welche Outcomes untersucht werden sollen. So untersuchte z. B. die *SAPALDIA Kohorte* (Swiss study on Air Pollution And Lung Disease in Adults) den Einfluss der Luftverschmutzung auf die Gesundheit der Atemwege und des Herz-Kreislauf-Systems. Es gibt jedoch auch Kohortenstudien, bei denen es keine vorformulierte Forschungsfrage gibt. In solchen Kohorten werden eine Vielzahl potenziell interessanter Expositionen sowie möglicher Confounder (s. Kap. 2.1.8) gemessen. Gleichzeitig werden unterschiedliche Outcomes erfasst. Ein Beispiel dafür ist die *Nationale Kohorte* in Deutschland. Hier sollen 200.000 Menschen über einen Zeitraum von mindestens 10 bis 20 Jahren beobachtet werden. Ein solches Design erlaubt es den EpidemiologInnen, im Verlauf der Studie eine Vielzahl von Forschungsfragen zu generieren. Sie identifizieren dann jeweils exponierte und nicht exponierte Untergruppen in der Kohorte und ermitteln, ob der für die jeweilige Forschungsfrage interessierende Outcome eintritt.

### Fall-Kontroll-Studien

Fall-Kontroll-Studien sind retrospektiv, also zeitlich gesehen „zurückschauend“ (s. Abb. 2.2). Der Outcome ist bereits eingetreten, es wird nun der Expositionsstatus bei Personen mit Outcome (*Fälle*) und ohne Outcome (*Kontrollen*) erfragt und miteinander verglichen. Als Assoziationsmaß in Fall-Kontroll-Studien dient die *Odds Ratio* (OR), ein Näherungswert für das Relative Risiko (RR). Inzidenzraten lassen sich in Fall-Kontroll-Studien nicht berechnen, da nicht die gesamte Population, aus der die Fälle stammen, sondern nur ausgewählte Kontrollen in die Studie aufgenommen werden.



**Abb. 2.2:** Schema einer Fall-Kontroll-Studie.

Fall-Kontroll-Studien sind besonders geeignet, wenn

- mehrere Expositionen untersucht werden sollen, die möglicherweise mit einem bestimmten Outcome assoziiert sind (z. B. verschiedene Expositionen, die das Risiko für Herz-Kreislauf-Erkrankungen erhöhen)
- der Outcome selten ist

Wollte man die Ursachen einer seltenen Krebserkrankung in einer Kohortenstudie untersuchen, so müsste man hunderttausende von Personen über einen längeren Zeitraum beobachten, um genügend inzidente (d. h. neu aufgetretene) Fälle zu finden. Einfacher und effizienter ist es, die Fälle, die über mehrere Jahre in einer großen Region aufgetreten sind, etwa mithilfe eines Krebsregisters zu identifizieren und dann zu befragen. Allerdings hängt die Qualität solcher Daten vom Erinnerungsvermögen der Fälle und der Kontrollen ab. Da die Rückfragezeiträume oft Jahre oder Jahrzehnte umfassen, kann es dabei zu Ungenauigkeiten kommen. Erinnern sich Fälle besser als Kontrollen – etwa, weil sie intensiv über mögliche Ursachen ihrer Krankheit nachdenken –, so kann das zu Verzerrungen führen (*Recall Bias*, s. Kap. 2.1.8).

Eine weitere Schwierigkeit bei Fall-Kontroll-Studien ist die Auswahl einer geeigneten Gruppe von Kontrollen. Idealerweise sollten sie als Zufallsstichprobe aus der gleichen Bevölkerung rekrutiert werden, aus der auch die Fälle stammen. Das ist mit einem hohen organisatorischen Aufwand verbunden, zudem ist die Teilnahmebereitschaft oft nur gering, was wiederum zu Verzerrungen führen kann (*Non-Response Bias*, s. Kap. 2.1.8). Eine mögliche Alternative sind Kontrollen aus einem Krankenhaus. Sie werden aus Abteilungen oder Stationen gezogen, in denen PatientInnen liegen, die das gesuchte Outcome nicht haben. In einer Studie zu oralen Kontrazeptiva („Pille“) und Brustkrebs könnten das beispielsweise Patientinnen einer orthopädischen Station sein. Es gibt aber immer wieder Belege dafür, dass sich KrankenhauspatientInnen hinsichtlich potenziell interessierender Expositionen wie Rauchen oder Alkoholkonsum von der Allgemeinbevölkerung unterscheiden. Sie sind also für diese nicht wirklich repräsentativ. Daraus kann wiederum eine Verzerrung aufgrund eines *Selektionsbias* (s. Kap. 2.1.8) resultieren.

Eine Fall-Kontroll-Studie kann nur dann durchgeführt werden, wenn die Exposition in der Bevölkerung nicht allzu selten ist, und wenn der Expositionsstatus von den Befragten erinnert werden kann. Letzteres ist nicht immer der Fall: Beim EHEC-Ausbruch 2011 (s. Kap. 8.2.3) in Deutschland waren Sprossen der Überträger der Erkrankung. Die ersten Fall-Kontroll-Studien im Rahmen der Ausbruchsuntersuchung konnten dies allerdings nicht nachweisen. Die EHEC-PatientInnen erinnerten sich nicht daran, dass sie Sprossen gegessen hatten. Die Sprossen waren meist Dekoration oder unauffällige Beigabe zu Salaten. Gut erinnert – und in der Folge fälschlich beschuldigt – wurden lediglich die optisch viel auffallenderen Gurken und Tomaten im Salat, die jedoch für die Übertragung gar nicht verantwortlich waren. Das Erfragen der Exposition und die Auswahl einer geeigneten Kontrollgruppe sind also eine besonderen Herausforderungen bei Fall-Kontroll-Studien.

### 2.1.6 Klinische Studien

Klinische Studien werden mit PatientInnen oder gesunden ProbandInnen durchgeführt, um Medikamente, bestimmte Behandlungsformen oder andere medizinische Interven-

tionen auf ihre Wirksamkeit und Sicherheit zu überprüfen. Klinische Studien sind meist experimentelle Studien. Anders als bei den in Kap. 2.1.5 vorgestellten beobachtenden Studientypen teilen ForscherInnen hier einer Studiengruppe eine bestimmte Intervention zu. Die klinische Prüfung verläuft in vier Phasen (hier dargestellt am Beispiel einer Medikamentenstudie):

- *Phase I*  
Studien an einer kleinen Zahl gesunder ProbandInnen (10–50), um die Wirkungen eines Medikaments am potentiellen Wirkort (Organsystem) zu untersuchen. Im Blickpunkt stehen darüber hinaus die Pharmakokinetik, Verträglichkeit und Sicherheit der Substanz.
- *Phase II*  
Erprobung an einer größeren Zahl von PatientInnen (bis ca. 100), um erste Hinweise auf die Wirksamkeit des Präparats und die notwendige Dosierung zu erhalten.
- *Phase III*  
Studie der therapeutischen Wirksamkeit an einer großen Zahl von PatientInnen (einige hundert bis Tausende). Gelingt dieser Wirkungsnachweis, wird in der Regel die Marktzulassung für das Medikament beantragt. Bei den Phase-III-Studien handelt es sich um randomisierte, kontrollierte Studien (*Randomized Controlled Trials*, RCT).
- *Phase IV*  
Studien nach der Zulassung des Medikaments, um unter den behandelten PatientInnen mögliche unerwünschte Arzneimittelwirkungen erkennen zu können. Phase-IV-Studien sind *beobachtende Studien*, da hier die Zuteilung zu der mit einem bestimmten Medikament behandelten Gruppe in der Routineversorgung der Arztpraxis und nicht kontrolliert durch die ForscherInnen vorgenommen wird.

### Randomisierte, kontrollierte Studien

In Phase-III-Studien wird die Wirksamkeit eines neuen Medikaments (*Verum*) mit der eines alten Medikaments oder der eines Scheinmedikaments (*Placebo*) verglichen. Die Gruppe, die das alte Medikament oder das Placebo erhält, wird auch als Kontrollgruppe oder Kontrollarm bezeichnet, die Verum-Gruppe auch als Interventionsgruppe oder Interventionsarm. Bei diesem Design kann es zu Verzerrungen kommen, wenn sich die PatientInnen der Interventions- und der Kontrollgruppe in Eigenschaften unterscheiden, die Einfluss auf das Behandlungsergebnis haben. Dies wäre z.B. der Fall, wenn ein Altersunterschied zwischen den Gruppen bestünde und sich die Heilungsaussichten mit zunehmendem Alter verschlechtern (*Confounding*, s. Kap. 2.1.8). Auch wäre es denkbar, dass ÄrztInnen besonders schwer erkrankte PatientInnen bevorzugt in die Verum-Gruppe aufnehmen – in der Hoffnung, dass die PatientInnen dort besonders effektiv behandelt werden. In der Verum-Gruppe befänden sich nun mehr schwerer Erkrankte, die leichteren Fälle verblieben in der Kontrollgruppe. Als Folge davon würde die Wirksamkeit des neuen Medikaments unterschätzt.

Eine ungleiche Verteilung prognostischer Faktoren zwischen beiden Gruppen muss also vermieden werden. Um dieses Ziel zu erreichen, werden die StudienteilnehmerInnen zufallsgesteuert entweder der Interventionsgruppe oder der Kontrollgruppe zugewiesen. Ziel einer solchen *Randomisierung* ist es, Strukturgleichheit herzustellen: Mögliche Störfaktoren werden mit gleicher Wahrscheinlichkeit auf die Interventions- und



die Kontrollgruppe verteilt. Bei größeren Studien mit mehreren hundert StudienteilnehmerInnen pro Arm kann davon ausgegangen werden, dass sich bekannte (und unbekannte) prognostische Faktoren gleichmäßig auf die Studiengruppen verteilen, womit eine verzerrende Wirkung verhindert wird.

**Randomisierung:** Von einer Randomisierung kann nur dann gesprochen werden, wenn die Zuteilung wirklich zufallsgesteuert erfolgt (z. B. durch computergenerierte Randomisierungslisten). Andere denkbare Verteilungsverfahren beinhalten die Möglichkeit einer Verzerrung. Wenn etwa alle PatientInnen, die montags in ein Krankenhaus aufgenommen werden, der Interventionsgruppe zugeteilt werden, während die Dienstags-PatientInnen in die Kontrollgruppe kommen, ist eine Strukturgleichheit beider Gruppen nicht gewährleistet. Es ist dann durchaus denkbar, dass montags (d. h. nach dem Wochenende) viele besonders schwere Fälle aufgenommen werden. Ein weiteres Problem ergibt sich aus der Vorhersehbarkeit der Zuordnung: Die für die Aufnahme der PatientInnen verantwortlichen Personen können die Zuordnung beeinflussen, indem sie z. B. diejenigen PatientInnen montags in die Studie aufnehmen, die ihrer Ansicht nach stärker von einer Therapie mit dem neuen Medikament profitieren. Die Zuordnung muss deshalb auch bei Verwendung von Randomisierungslisten immer verdeckt durch eine außenstehende Stelle vorgenommen werden, die unabhängig vom direkt in die Studie involvierten Personal ist (*Concealment of Allocation*).

Ist die Zahl der StudienteilnehmerInnen gering, wählt man in der Regel eine *stratifizierte Randomisierung*. Hierdurch können wichtige Merkmale der PatientInnen gleichmäßig auf die Interventions- und die Kontrollgruppe verteilt werden, ohne das Prinzip der Randomisierung zu unterlaufen. Dies kann z. B. dann angebracht sein, wenn vermutet wird, dass ein Impfstoff bei Männern und Frauen unterschiedlich wirkt. Die PatientInnen werden dabei zunächst in Geschlechtergruppen aufgeteilt (stratifiziert). Anschließend wird die Randomisierung jeweils innerhalb dieser Geschlechtergruppen vorgenommen. Mit der *Blockrandomisierung* erreicht man, dass jeweils gleich viele PatientInnen den vorhandenen Studiengruppen zugeordnet werden. Dabei wird für jeden PatientInnen-Block (z. B. für jeweils 8 Personen) zufallsgesteuert festgelegt, in welcher Reihenfolge sie der Interventions- bzw. der Kontrollgruppe zugeteilt werden. Hierbei muss das zuvor definierte Verhältnis eingehalten werden (z. B. vier PatientInnen in jeder der beiden Gruppen). So ist sichergestellt, dass die entstehenden Gruppen zufallsverteilt gleich groß sind.

**Verblindung:** Zu einer Verzerrung der Studienergebnisse kann es auch dann kommen, wenn z. B. ÄrztInnen das Behandlungsergebnis unbewusst als besonders positiv beurteilen, weil sie wissen, dass der/die untersuchte PatientIn das Verum (und nicht das Placebo) erhalten hat. UntersucherInnen können also durch die Kenntnis der Studienhypothese und der Gruppenzugehörigkeit in ihrer Beurteilung des Behandlungsergebnisses beeinflusst werden. Um eine solche Verzerrung zu vermeiden, wird eine so genannte Verblindung durchgeführt (*Blinding*). Hierbei wissen weder PatientInnen noch UntersucherInnen, ob das Verum oder ein Placebo gegeben wurde. Voraussetzung dafür ist, dass sich Verum- und Placebopräparat optisch gleichen. In der Regel werden heute Phase-III-Studien als randomisierte, kontrollierte Studien (RCT) durchgeführt, bei denen sowohl PatientInnen als auch untersuchende ÄrztInnen verblindet wurden (*Doppelblind-Studie*). In einigen Studien werden auch diejenigen Personen verblindet, die die Daten bearbeiten. Für sie ist zunächst nicht erkennbar, welche Gruppe die Verum-

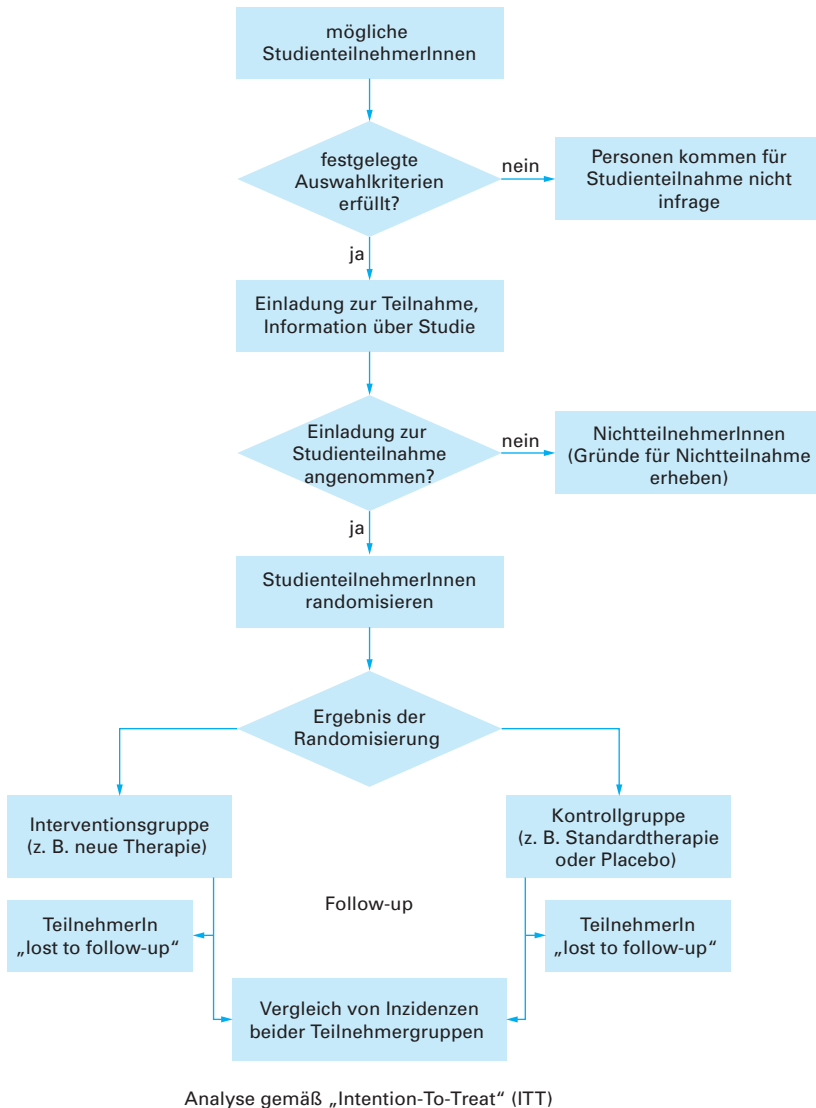
und welche die Placebo-Gruppe ist. Die Informationen hierzu werden von der Studienleitung unter Verschluss gehalten. Erst nach der Durchführung der Analyse wird offengelegt, welche Gruppe welches Präparat erhalten hat. Eine Verblindung aller Beteiligten ist nicht immer möglich (z. B. beim Vergleich von invasiven mit konservativen Therapien). Von der Verblindung zu unterscheiden ist die oben beschriebene *verdeckte Zuordnung* (Concealment of Allocation) bei der Randomisierung. Sie hilft Ungleichheiten der Studiengruppen bei der Randomisierung zu vermeiden und ist immer machbar.

**Studienablauf:** Bei der Planung einer RCT ist in einem ersten Schritt zu klären, welche PatientInnen die Voraussetzungen für die Teilnahme an der Studie erfüllen. Dazu müssen klare Ein- und Ausschlusskriterien definiert sein. Anschließend werden die potentiellen StudienteilnehmerInnen über das Ziel der Studie, die Randomisierung und die Verblindung aufgeklärt. Wenn sie ihr Einverständnis zur Teilnahme an der Studie gegeben haben (*Informed Consent*), werden sie in die Studie aufgenommen. Erst danach findet die Randomisierung statt. Die PatientInnen erhalten dann entweder das Verum- oder das Placebopräparat. Abb. 2.3 zeigt schematisch den Ablauf einer solchen Studie.

Die Auswertung der Daten erfolgt ähnlich wie bei einer Kohortenstudie. Es können z. B. die kumulativen Inzidenzen für Outcomes oder für das Auftreten von Nebenwirkungen in beiden Gruppen verglichen werden. Das Ergebnis wird dann als relatives Risiko angegeben. Auch in diesem Stadium einer klinischen Studie können sich Verzerrungen ergeben, etwa wenn PatientInnen nicht in der Behandlungsgruppe verbleiben, der sie durch die Randomisierung zugewiesen wurden. So wäre es denkbar, dass ein behandelnder Arzt eine Patientin bei einem besonders schweren Verlauf und fehlendem Hinweis auf Besserung der anderen Gruppe zuweist – in der Hoffnung, es könne sich dabei um die Verum-Gruppe mit einem wirksamen Medikament handeln. Auf diese Weise würde sich die Zahl der PatientInnen mit schwererem Verlauf in einer der beiden Gruppen erhöhen, was das Ergebnis verzerren würde. Um dies zu vermeiden, werden alle PatientInnen entsprechend ihrer ursprünglichen Zuordnung ausgewertet, unabhängig davon, welche Behandlung sie tatsächlich erhalten haben. Diese Analyse aufgrund der beabsichtigten Therapie wird mit *Intention-to-Treat* (ITT) bezeichnet und sollte in klinischen Studien Standard sein. Probleme entstehen auch dann, wenn PatientInnen ihre Teilnahme an der Studie aufkündigen oder der Kontakt zu den StudienteilnehmerInnen abreißt (*Loss to follow up*, LTFU).

**Fallzahlkalkulation (Sample Size Calculation):** Vor dem Beginn einer klinischen Studie muss zunächst die erforderliche Größe der Studie festgelegt werden. Mit Unterstützung eines Statistikers wird hierzu eine Fallzahlkalkulation vorgenommen. Zunächst ist zu überlegen, wie groß der Unterschied zwischen der Verum- und Kontrollgruppe mindestens sein muss, um klinisch relevant zu sein. Anschließend wird das *Signifikanzniveau* (s. Kap. 2.3.8) festgelegt – in der Regel entscheidet man sich für ein Signifikanzniveau von 5 %. Das bedeutet, dass die Wahrscheinlichkeit, diesen Unterschied (oder einen größeren Unterschied) zu beobachten, wenn in Wirklichkeit *kein* Unterschied besteht, kleiner als 5 % sein soll. Weiter muss die so genannte *Power* der Studie festgelegt werden. Die Power gibt die Wahrscheinlichkeit an, dass die Studie den Unterschied, wenn er wirklich besteht, als statistisch signifikant aufdeckt. Meist entscheidet man sich für eine Power von 80 %.

Wird eine klinische Studie mit einer zu kleinen Fallzahl durchgeführt, so sind die Ergebnisse statistisch nicht aussagekräftig, denn die Power der Studie reicht nicht aus,



**Abb. 2.3:** Schematischer Ablauf einer randomisierten, kontrollierten Studie (RCT).

um einen möglicherweise bestehenden Unterschied aufzuzeigen. In diesem Fall wird das 95%-Vertrauensintervall (siehe Kap. 2.3.6) des Relativen Risikos oder der Relativen Rate die 1 mit einschließen, d. h. die Ergebnisse für das Verumpräparat sind im Vergleich zum Placebo oder dem alten Medikament sowohl mit einer besseren als auch mit einer schlechteren Wirkung vereinbar. Die Studie hat ihren Zweck dann nicht erfüllt. Leider sind viele klinische Studien in der Tat zu klein, um klinisch relevante Unterschiede aufzuzeigen oder auszuschließen. In dieser Situation können *Meta-Analysen* mehrerer Studien sinnvoll sein (siehe Kap. 2.1.7).

**Relative und absolute Risikoreduktion:** Bei der Analyse der Daten einer randomisierten, kontrollierten Studie werden zunächst die *absoluten Risiken* oder kumulative Inzidenzraten (s. Kap. 2.1.2) in den beiden Therapiegruppen berechnet. Die *relative Risikoreduktion* (RRR) gibt an, um welchen Prozentsatz der Einsatz der Verum-Therapie das Ergebnis gegenüber der Kontrollgruppe verändert. Die relative Risikoreduktion berechnet sich aus der kumulativen Inzidenz in den beiden Studienarmen. Durch die Multiplikation mit 100 wird die relative Risikoreduktion in Prozent angegeben.

$$\text{RRR} = \frac{\text{Kumulative Inzidenz}_{\text{Kontrollgruppe}} - \text{Kumulative Inzidenz}_{\text{Interventionsgruppe}}}{\text{Kumulative Inzidenz}_{\text{Kontrollgruppe}}} \cdot 100$$

Die relative Risikoreduktion ist wenig aussagekräftig, wenn der untersuchte Outcome sehr selten ist. Eine Senkung einer sehr geringen Wahrscheinlichkeit um einen bestimmten Prozentsatz ist möglicherweise klinisch auch nicht relevant. Ein besseres Maß ist hier die *absolute Risikoreduktion* (ARR). Sie errechnet sich aus der Differenz der Risiken bzw. Inzidenzraten in beiden Therapiegruppen.

$$\text{ARR} = \text{Kumulative Inzidenz}_{\text{Kontrollgruppe}} - \text{Kumulative Inzidenz}_{\text{Interventionsgruppe}}$$

Ist die ARR größer als 0, so wirkt die Verum-Therapie besser als die Therapie, die in der Kontrollgruppe eingesetzt wurde (altes Medikament oder Placebo). Wirkt die neue Therapie jedoch nicht besser oder sogar schlechter als die Kontrollmedikation, ist die absolute Risikoreduktion gleich oder kleiner als 0. Da die ARR statt einer prozentualen Veränderung die tatsächliche Inzidenz angibt, ist die klinische Relevanz eines so angegebenen Ergebnisses leichter zu beurteilen.

**Number Needed to Treat/Number Needed to Harm:** Ein gebräuchliches Maß, um die Wirksamkeit eines neuen Medikaments zu beschreiben, ist die *Number Needed to Treat* (NNT). Sie gibt an, wie viele PatientInnen mit dem neuen anstatt dem alten Medikament behandelt werden müssen, um einen zusätzlichen Behandlungserfolg zu erzielen. Die NNT wird auch als „Anzahl der erforderlichen Behandlungen“ bezeichnet. Sie berechnet sich folgendermaßen:

$$\text{NNT} = \frac{1}{\text{ARR}} = \frac{1}{\text{Kumulative Inzidenz}_{\text{Kontrollgruppe}} - \text{Kumulative Inzidenz}_{\text{Interventionsgruppe}}}$$

Eine weitere, wichtige Messgröße ist die *Number Needed to Harm* (NNH). Sie gibt die Zahl der PatientInnen an, die mit der neuen Therapie behandelt werden müssen, bis im Vergleich zur Kontrollgruppe bei einem/r PatientIn eine zusätzliche unerwünschte Wirkung auftritt. Die NNH zeigt damit, wie häufig unerwünschte, durch das Verum-Präparat hervorgerufene Wirkungen (= Nebenwirkungen) sind. Ihre Berechnung erfolgt entsprechend dem Vorgehen bei der NNT.

Die Berechnung von RRR, ARR und NNT werden im Folgenden an einem Beispiel illustriert. In einer Studie zur Wirksamkeit eines neuen Cholesterinsenkers erhielten 2.275 PatientInnen das neue Medikament und 2.133 ein Placebo. Im Laufe des Follow-ups verstarben 182 PatientInnen in der Interventionsgruppe (Medikament) und 256 PatientInnen in der Kontrollgruppe (Placebo). Das Sterberisiko in der Interventionsgruppe betrug daher  $182/2275 = 0,08$  (oder 8 %), das in der Kontrollgruppe  $256/2133 = 0,12$  (oder 12 %). Die relative Risikoreduktion (RRR), die absolute Risikoreduktion (ARR) und

die Number Needed to Treat (NNT) lassen sich nun entsprechend den oben angegebenen Formeln wie folgt berechnen:

$$\text{RRR: } (0,12 - 0,08) / 0,12 \cdot 100 = 33,3 \%$$

$$\text{ARR: } (0,12 - 0,08) = 0,04 \text{ (oder } 4 \%)$$

$$\text{NNT: } 1 / 0,04 = 25.$$

Der Einsatz des neuen Cholesterinsenkers reduzierte damit die Sterblichkeit gegenüber der Kontrollgruppe um ca. 33 %, d.h. in der Interventionsgruppe starben ein Drittel weniger PatientInnen als in der Kontrollgruppe. Absolut betrachtet konnte das Medikament die Sterblichkeit jedoch nur um 4 % senken. Um einen Todesfall zu verhindern, müssen nun 25 PatientInnen über einen Zeitraum von durchschnittlich 5,4 Jahren mit dem neuen Medikament behandelt werden.

### Ethische Aspekte

Die ethischen Prinzipien, die in einer klinischen Studie berücksichtigt werden müssen, sind in der *Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects* (s. Internet-Ressourcen) aufgelistet. Diese Deklaration fordert, das genaue Versuchsprotokoll einer Ethikkommission vorzulegen. Die Ethikkommission prüft, ob die geplante Studie ethisch vertretbar und zulässig ist. Hierfür wird u.a. beurteilt, ob die möglichen Nachteile, die mit einer Studienteilnahme verbunden sind, in einem vertretbaren Verhältnis zu den möglichen Vorteilen der Studie stehen (Nutzen-Risiko-Verhältnis). Weiterhin wird geprüft, ob Patienteninformation und Einwilligungserklärung umfassend und verständlich sind. Ethikkommissionen sind unabhängige Gremien und bestehen meist aus ÄrztInnen, Pflegefachkräften, StatistikerInnen, JuristInnen und MedizinethikerInnen oder TheologInnen.

Internationale Richtlinien der *guten klinischen Praxis* (s. Internetressourcen) definieren Standards, die die Durchführung einer klinischen Studie regeln. Die Einhaltung dieser Standards gewährleistet die Qualität der Studien und erleichtert Vergleiche zwischen verschiedenen Studien. Die Berichterstattung und Dokumentation zu randomisierten klinischen Studien sollte auf der Basis der *CONSORT-Richtlinien* (*Consolidated Standards of Reporting Trials Statement*, s. Internetressourcen) erfolgen. Zu einer solchen ordnungsgemäßen Berichterstattung gehört, dass die Ergebnisse *aller* randomisierten, kontrollierten Studien publik gemacht werden – auch wenn sie gezeigt haben, dass ein neues Präparat nicht wirksamer ist als ein altes Präparat oder ein Placebo. Eine solche Offenlegung aller Studienergebnisse ist besonders wichtig für die Durchführung von Meta-Analysen (s. Kap. 2.1.7), weil auf diese Weise eine Verzerrung der Resultate durch einen *Publikationsbias* (eine Form von *Selektionsbias*, s. Kap. 2.1.8) verhindert werden kann.

Heute sind viele wissenschaftliche Zeitschriften nur dann bereit, eine randomisierte, kontrollierte Studie zu veröffentlichen, wenn die Studie zuvor in ein spezielles Register aufgenommen wurde (z. B. *Deutsches Register Klinischer Studien* oder *EU Clinical Trials Register*). Die von der Weltgesundheitsorganisation WHO unterstützten und koordinierten Register sollen verhindern, dass negative Ergebnisse bei klinischen Studien unter Verschluss bleiben.

## Randomisierte, kontrollierte Studien bei komplexen Interventionen

Randomisierte, kontrollierte Studien kommen auch außerhalb der klinischen Forschung zum Einsatz, etwa um die Wirksamkeit von Maßnahmen der Verhaltensprävention (z. B. der Raucherentwöhnung, s. Kap. 4.6), von Screening-Programmen (s. Kap. 4.5.3) oder Versorgungsmodellen (z. B. *Managed Care*, s. Kap. 3.2.2) zu evaluieren. In der Tat wird zunehmend gefordert, auch Public-Health-Interventionen in randomisierten, kontrollierten Studien zu erproben und damit eine stärkere Evidenzbasierung von Public Health zu schaffen (*Evidence-based Public Health*, EbPH).

Auch bei diesen randomisierten, kontrollierten Studien teilen ForscherInnen einer Studiengruppe eine bestimmte Intervention zu, und auch hier gibt es eine Kontrollgruppe. Die Interventionen sind im Gegensatz zu klinischen Medikamentenstudien jedoch oft komplexer und Placebo-Kontrollen in der Regel nicht möglich. Dies erschwert die Verblindung der StudienteilnehmerInnen oder macht sie unmöglich. Ein weiteres Problem ist die *Kontamination* des Interventionseffektes, die dadurch entsteht, dass sich StudienteilnehmerInnen aus Interventions- und Kontrollgruppe austauschen. Die Interventionsmaßnahmen gelangen so in die Kontrollgruppe. Um dies zu verhindern, werden oft nicht einzelne Personen sondern Arztpraxen, Pflegestationen, Dörfer oder Versorgungsregionen randomisiert. Ein weiterer Grund für eine solche *Cluster-Randomisierung* liegt darin, dass sich Interventionen oft nicht auf der individuellen Ebene umsetzen lassen.

### 2.1.7 Systematische Übersichten und Meta-Analysen

Angesichts der großen Anzahl publizierter Einzelstudien sind gute Übersichten, die Auskunft über die vorhandene Evidenzlage geben, für die Entscheidungsfindung in der klinischen Medizin und in der Gesundheitspolitik besonders wichtig. Wir unterscheiden in diesem Zusammenhang zwischen narrativen und systematischen Übersichtsarbeiten einerseits und Meta-Analysen andererseits.

**Narrative Übersichten** fassen die Ergebnisse wichtiger Studien informell zusammen und kommentieren sie. Die Auswahl und Beurteilung der Studien ist subjektiv und nicht immer umfassend. Die Schlussfolgerungen, die die Autoren ziehen, spiegeln nicht zuletzt auch ihre persönliche Meinung wider. Gerade aus diesem Grund haben narrative Übersichtsarbeiten jedoch auch weiterhin ihren Platz in der Literatur.

Im Unterschied zu narrativen Übersichtsarbeiten zeichnen sich **systematische Übersichtsarbeiten** durch eine klar definierte Fragestellung und ein reproduzierbares, wissenschaftliches Vorgehen aus. Es handelt sich hierbei um Studien über Studien. Nicht Personen werden untersucht, sondern alle Studien, die die zuvor festgelegten Einschlusskriterien erfüllen. In systematischen Übersichtsarbeiten wird die methodologische Qualität der in die Arbeit aufgenommenen Studien nach festgelegten Kriterien beurteilt, die Ergebnisse werden standardisiert erfasst. Ein- und Ausschlusskriterien, die Kriterien für die Studienqualität, Outcomes und Interventionen werden ebenso wie die Literatursuche ausführlich beschrieben. Auch mögliche Verzerrungen durch *Publikationsbias* (s. Kap. 2.1.6 und Kap. 2.1.8) oder eine mangelhafte Qualität der Studien werden diskutiert. Auf diese Weise werden die Schlussfolgerungen der Autoren leicht nachvollziehbar und reproduzierbar. Die *Cochrane Collaboration* ist ein internationales Netzwerk von WissenschaftlerInnen, ÄrztInnen und PatientInnen, das qualitativ hoch-



stehende systematische Übersichtsarbeiten verfasst. Ihr Ziel ist es, dadurch evidenzbasiertes Handeln in der medizinischen Versorgung und in Public Health zu fördern. Ihre Übersichtsarbeiten werden über die *Cochrane Library* zugänglich gemacht.

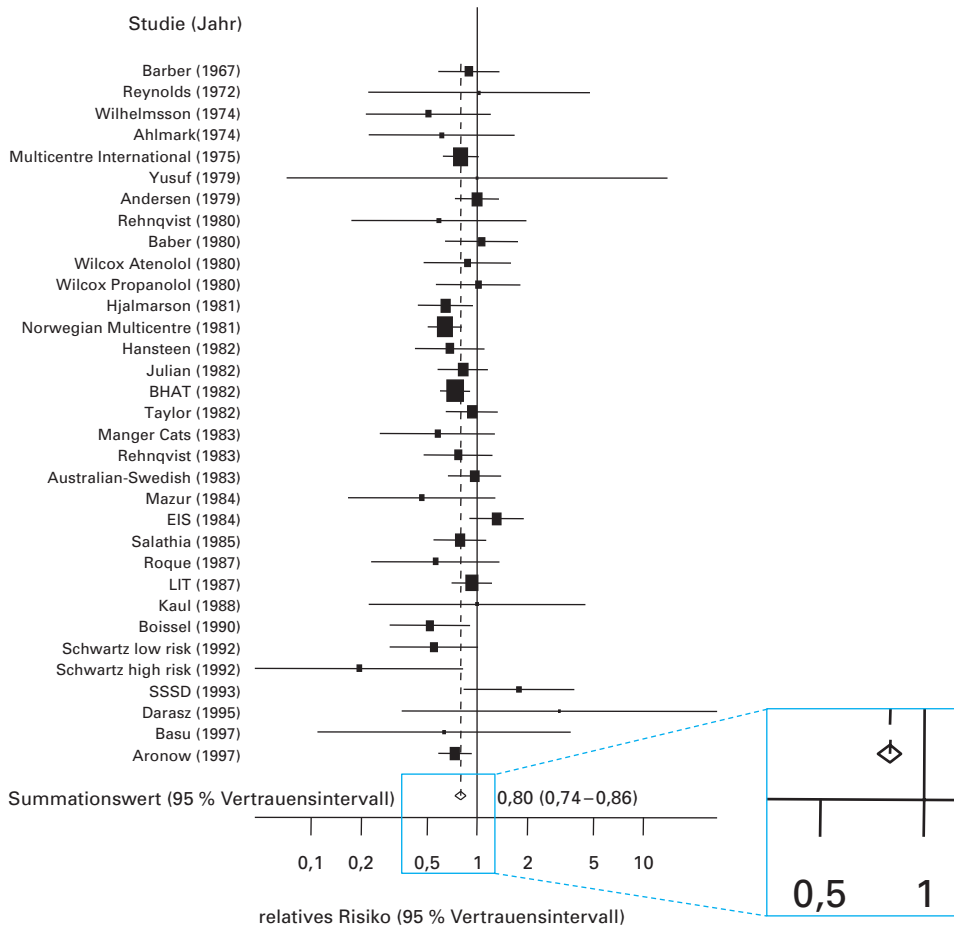
Systematische Reviews enthalten oft eine oder mehrere **Meta-Analysen**. In Meta-Analysen werden die Ergebnisse der verschiedenen Studien statistisch zusammengefasst, um präzisere und allgemeingültigere Angaben über die Wirksamkeit von Interventionen zu erhalten. Einzelne Studien sind oft nicht groß genug, um kleinere, aber klinisch bedeutende Unterschiede sicher zu erfassen. Zudem sind die Ergebnisse von Einzelstudien oft nur auf eine relativ eng umschriebene Population anwendbar. Im Zentrum der Meta-Analyse steht die Berechnung eines *Summationswertes* mit einem Vertrauensintervall. Beim Summationswert handelt es sich um einen gewichteten Mittelwert, der große Studien stärker gewichtet als kleinere Studien. Der Summationswert wird zusammen mit den Ergebnissen der einzelnen Studien in einem *Forest Plot* dargestellt.

Abb. 2.4 zeigt den Forest Plot einer Meta-Analyse von randomisierten klinischen Studien, die die sekundärpräventive Wirkung von verschiedenen Betablockern auf die Mortalität bei Personen nach einem Herzinfarkt untersuchten. Die schwarzen Rechtecke repräsentieren die gefundene Wirkung auf das Sterblichkeitsrisiko in den einzelnen Studien. Die Flächen der Rechtecke zeigen dabei das Gewicht der Studien in der Meta-Analyse, die horizontalen Linien die 95%-Vertrauensintervalle. Die vertikale Linie stellt das relative Risiko von 1 dar. Hier gibt es also keinen Unterschied in der Wirkung zwischen Betablocker- und Kontrollgruppen. In Studien, die links von dieser Linie liegen, war die Sterblichkeit in der Betablocker-Gruppe niedriger als in der Kontrollgruppe, während auf der rechten Seite der Linie die Kontrollgruppen besser abschnitten. Die gestrichelte Linie zeigt den Summationswert aus der Meta-Analyse an, das Diamant-Zeichen unterhalb der aufgelisteten Einzelstudien veranschaulicht das 95%-Vertrauensintervall des Summationswertes (s. vergrößerter Ausschnitt in Abb. 2.5). *Achtung:* In Forest Plots werden logarithmische Skalen verwendet, sodass 0,5 und 2 den gleichen Abstand zu 1 aufweisen und die Vertrauensintervalle damit symmetrisch sind!

In den meisten 95%-Vertrauensintervallen unserer Abbildung ist die 1 enthalten. Die Mehrzahl der Studien zeigt somit, dass Betablocker keine statistisch signifikante sekundärpräventive Wirkung auf die Mortalität bei Personen nach einem Herzinfarkt haben ( $p > 0.05$ ). Durch die Meta-Analyse wird hingegen deutlich, dass die Einnahme von Betablockern in dieser Situation zu einer Reduktion der Sterblichkeit um 20 % führt. Das errechnete enge Vertrauensintervall schließt die 1 hier nicht ein (relatives Risiko = 0,80; 95%-Vertrauensintervall 0,74–0,86). Anhand der gestrichelten Linie ist eine visuelle Beurteilung der Variabilität bzw. Heterogenität der Ergebnisse der Studien möglich. In unserem Beispiel ist diese recht klein. Die Ergebnisse der Studien liegen nahe beieinander, die 95%-Vertrauensintervalle schließen den Summationswert aus der Meta-Analyse mit ein. Es handelt sich somit um eine homogene Situation. Dies legt die Schlussfolgerung nahe, dass der Wirkung der verschiedenen Betablocker ein Klasseneffekt<sup>6</sup> zugrunde liegt, der in unterschiedlichen Patientenpopulationen zum Tragen kommt und deshalb eine hohe externe Validität (Generalisierbarkeit) aufweist.

<sup>6</sup> *Klasseneffekt:* Ein Effekt, der bei einer Wirkstoff-Klasse und nicht nur bei einem einzelnen Medikament auftritt.





**Abb. 2.4:** Meta-Analyse von randomisierten, klinischen Studien zur sekundärpräventiven Wirksamkeit einer Therapie mit Betablockern bei PatientInnen nach Myokardinfarkt.

Meta-Analysen können die Wirksamkeit von Medikamenten und anderen Interventionen oft früher und überzeugender nachweisen als einzelne kleinere Studien. In diesem Fall wäre der Nutzen der Betablocker bereits Anfang der 1980er Jahre nachweisbar gewesen. Die Methode sollte jedoch keinesfalls unkritisch angewendet werden. Grundsätzlich kann die Qualität einer Meta-Analyse nicht besser sein als diejenige der Einzelstudien (*Garbage in – Garbage out*<sup>7</sup>). Auch wird der Heterogenität der Ergebnisse oft nicht genügend Beachtung geschenkt. Wenn stark voneinander abweichende Resultate vorliegen, ist eine statistische Kombination dieser Ergebnisse selten sinnvoll. Schließlich kann ein *Publicationsbias* in den ausgewählten Studien die Ergebnisse von Meta-Analysen verzerren.

<sup>7</sup> *Garbage in – Garbage out* (engl.): Wo man Müll hineinsteckt, kommt auch Müll heraus.

### 2.1.8 Mögliche Fehlerquellen in epidemiologischen Untersuchungen

Bevor beurteilt werden kann, ob eine ermittelte *Assoziation* zwischen einer Exposition und einem Outcome kausal ist, d. h. ob hier eine Ursache-Wirkungs-Beziehung besteht, muss zunächst untersucht werden, inwiefern mögliche Fehler und Verzerrungen den wahren Zusammenhang zwischen beiden verschleiern.

#### Zufällige Fehler

Ein zufälliger Fehler liegt dann vor, wenn der Wert einer gemessenen Variablen in einer Stichprobe rein zufallsbedingt um den tatsächlichen (wahren) Wert der Variablen in der Grundgesamtheit streut. Eine häufige Quelle von zufälligen Fehlern sind ungenaue Messungen. Ist die Stichprobe groß genug, fallen Messungenauigkeiten in der Regel nicht ins Gewicht. Die einzelnen Messwerte schwanken dann zwar um den wahren Wert und die *Reliabilität* der Messung ist gering, denn bei jeder Messung wird ein anderer Wert ermittelt. Im Durchschnitt nähern sich die einzelnen Messwerte aber dem wahren Wert an, die *Validität* der Messung ist daher hoch (s. Kap. 2.1.4).

Um mit ausreichender Sicherheit auf die Grundgesamtheit schließen zu können, ist auch bei genauen Messungen eine ausreichend große Stichprobe nötig. Wenn die Stichprobe zu klein ist, können trotz genauer Messungen zufällige Fehler in Form von so genannten *Stichprobenfehlern* auftreten. Dies lässt sich leicht mit Hilfe eines Würfels veranschaulichen. Wird nur zehnmal gewürfelt, kann es passieren, dass manche Augenzahlen oft, andere vielleicht gar nicht gewürfelt werden, obwohl bei einem „fairen“ Würfel alle sechs Augenzahlen die gleiche Wahrscheinlichkeit haben, geworfen zu werden (nämlich ein Sechstel). Stichprobenfehler nehmen mit zunehmender Stichprobengröße ab. Wie groß die Stichprobe letztendlich sein muss, um den zufälligen Fehler auf einem akzeptablen und vor Studienbeginn definierten Niveau zu halten, ist von unterschiedlichen Faktoren abhängig und Gegenstand der *Fallzahlplanung* (s. Kap. 2.1.6).

#### Systematische Fehler

Von einem systematischen Fehler oder **Bias** ist in der Epidemiologie dann die Rede, wenn der Fehler nicht zufällig, sondern immer in der gleichen Weise auftritt. Ein Beispiel hierfür ist eine ungeeichte Waage, die das Gewicht einer Person immer um 5 Kilogramm zu hoch angibt. Eine solche Messung ist *reliabel*, denn bei jeder Messung wird der gleiche Wert gemessen. Ihre Validität ist jedoch gering, da auch mit zunehmender Anzahl von Messungen der korrekte (wahre) Wert nicht ermittelt werden kann (vgl. Kap. 2.1.4).

**Selektionsbias:** Ein Selektionsbias bezeichnet einen systematischen Auswahlfehler bei der Rekrutierung von StudienteilnehmerInnen oder bei deren Verbleib in einer Studie. Ein Beispiel für einen häufigen Auswahlfehler ist ein sogenannter *Non-Response-Bias*. Er kann auftreten, wenn sich die TeilnehmerInnen einer Studie von den Personen unterscheiden, die eine Studienteilnahme verweigern. Ein Grund hierfür ist, dass Eigenschaften, die mit der Teilnahmeverweigerung zusammenhängen, oft auch mit der Exposition und dem Outcome assoziiert sind. So konsumieren Teilnahmeverweigerer (so

genannte *Non-Responder*) im Vergleich zu TeilnehmerInnen oft mehr Alkohol und Tabak. Um den Grad der Verzerrung abzuschätzen, die durch einen Non-Response-Bias entsteht, müssen daher auch grundlegende Informationen über Non-Responder eingeholt werden. Eine spezielle Variante des Non-Response-Bias kann in der Beobachtungs-Phase (*Follow-up*) von Kohorten- und randomisierten kontrollierten Studien auftreten. Dies ist der Fall, wenn sich die Studienausfälle in der Gruppe der Exponierten und nicht Exponierten (bzw. der Interventions- und Kontrollgruppe) in Faktoren unterscheiden, die auch mit der Exposition oder dem Outcome in Zusammenhang stehen. Um diesen sogenannten *Loss-to-follow-up-Bias* zu vermeiden, müssen StudienteilnehmerInnen in Längsschnittstudien intensiv nachverfolgt werden (vgl. Abb. 2.3).

Ein Selektionsbias kann nicht nur StudienteilnehmerInnen, sondern auch wissenschaftliche Ergebnisse betreffen. WissenschaftlerInnen und angesehene wissenschaftliche Publikationsorgane neigen nämlich dazu, eher Ergebnisse zu veröffentlichen, die Zusammenhänge zwischen Expositionen und Outcomes aufzeigen. Studien, die (wider Erwarten) keine Assoziationen erkennen lassen, werden daher oft gar nicht oder in Zeitschriften veröffentlicht, die wenig Beachtung finden. Für Meta-Analysen stellt dieser so genannte *Publikationsbias* ein Problem dar, denn er kann zu einer Überschätzung des tatsächlichen Zusammenhangs zwischen Expositionen und Outcomes führen (s. Kap. 2.1.7).

**Informationsbias:** Ein Informationsbias (auch *Missklassifikation* genannt) bezeichnet einen systematischen Messfehler, der dazu führt, dass Personen im Hinblick auf Exposition, Outcome und weitere Einflussvariablen (*Kovariaten*) falsch klassifiziert werden. Exponierte werden z.B. fälschlicherweise den nicht Exponierten zugeordnet oder Kranke den Gesunden.

Die möglichen Folgen eines Informationsbias werden im Folgenden an der in Kap. 2.1.3 beschriebenen Fall-Kontroll-Studie zur Assoziation von Neuroleptikaeinnahme und venöser Thromboembolie (VTE) illustriert. Die AutorInnen ermittelten in ihrer Untersuchung eine Odds Ratio von 1,6. Wir nehmen nun an, dass sowohl bei den Fällen (VTE) als auch den Kontrollen (keine VTE) 5% aller Personen, die keine Neuroleptika einnahmen, dies bei der Befragung nicht korrekt angaben und damit fälschlicherweise der Gruppe der Exponierten zugeordnet wurden. Eine solche Missklassifikation, die unabhängig vom Outcome- oder Expositionsstatus ist, wird als *nicht-differenzielle Missklassifikation* bezeichnet. Die Vier-Felder-Tafel hierzu ist in Tab. 2.3 dargestellt.

**Tab. 2.3:** Vier-Felder-Tafel zum Zusammenhang von Neuroleptikaeinnahme und venöser Thromboembolie (VTE; Zahlen nach Parker et al. 2010, s. Tab. 2.2) mit einer hypothetischen fünfprozentigen *nicht-differenziellen Missklassifikation* (oben) und einer hypothetischen *differenziellen Missklassifikation* (unten).

- (1) Bei der nicht-differenziellen Missklassifikation wurden 5% aller Personen, die keine Neuroleptika einnahmen, fälschlicherweise der Gruppe der Exponierten zugeordnet.
- (2) Die differenzielle Missklassifikation beträgt bei den erkrankten Personen 10% (d.h. 10% aller Personen, die keine Neuroleptika einnahmen, wurden fälschlicherweise der Gruppe der Exponierten zugeordnet) und bei den nicht erkrankten Personen 5%.

**Nicht-differenzielle Missklassifikation**

		Outcome (VTE)				Summe	
		Ja		Nein			
Exposition (Neuroleptikaeinnahme)	Ja	3.296	a	b	8.989	a+b	12.285
	Nein	22.236	c	d	80.502	c+d	102.738
Summe		25.532	a+c	b+d	89.491	a+b+c+d	115.023

**Odds Ratio** =  $(3.296 \cdot 80.502) / (8.989 \cdot 22.236) = 1,33$

**Differenzielle Missklassifikation**

		Outcome (VTE)					
		Ja		Nein			
Exposition (Neuroleptikaeinnahme)	Ja	4.467	a	b	8.989	a+b	13.456
	Nein	21.065	c	d	80.502	c+d	101.567
Summe		25.532	a+c	b+d	89.491	a+b+c+d	115.023

**Odds Ratio** =  $(4.467 \cdot 80.502) / (8.989 \cdot 21.065) = 1,90$

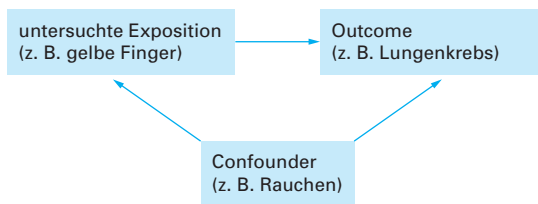
Die Odds Ratio ist hier mit  $(3.296 \cdot 80.502) / (8.989 \cdot 22.236) = 1,3$  geringer als zuvor. Wie in diesem Beispiel führt eine nicht-differenzielle Missklassifikation fast immer zu einer Unterschätzung des Zusammenhangs von Exposition und Outcome.

Dagegen können Missklassifikationen, die vom Outcome- oder Expositionsstatus abhängig sind (so genannte *differenzielle Missklassifikationen*), sowohl zu einer Unter- als auch zu einer Überschätzung der Assoziation von Exposition und Outcome führen. Zu einer differenziellen Missklassifikation kann es z. B. infolge eines unterschiedlichen Erinnerungsvermögens bei Fällen und Kontrollen kommen (*Recall-Bias*) – ein Problem, das häufig in Fall-Kontroll-Studien auftritt. Ein solcher Recall-Bias könnte beispielsweise in der genannten Fall-Kontroll-Studie dazu führen, dass die Missklassifikation bei den erkrankten Personen mit 10% höher ist als bei den nicht erkrankten Personen, wo sie nur 5% beträgt (Tab. 2.3). In diesem hypothetischen Fall einer differenziellen Missklassifikation beträgt die Odds Ratio 1,9. Die Assoziation wird daher im Vergleich zu den Originaldaten überschätzt.

## Confounding

Der Begriff Confounding bezeichnet eine Verzerrung, die durch den Einfluss einer oder mehrerer weiterer Einflussvariablen (so genannter *Confounder* oder *Störgrößen*) entsteht, sodass die Assoziation zwischen Exposition und Outcome über- oder unterschätzt wird.

Confounding liegt dann vor, wenn diese Einflussvariablen sowohl mit dem Outcome als auch mit der Exposition assoziiert sind. Ein Beispiel für Confounding ist die Assoziation von gelben Fingern (*Exposition*) und Lungenkrebs (*Outcome*). Menschen, die viel rauchen, haben bedingt durch das Kondensat des Zigarettenrauchs oft gelbe Finger. Rauchen ist aber auch ein bekannter Risikofaktor für Lungenkrebs, und zwar unabhängig davon, ob die RaucherInnen gelbe Finger haben oder nicht. Das Rauchen ist also im Hinblick auf den Zusammenhang zwischen gelben Fingern und Lungenkrebs ein Confounder. Wird dieser Confounder nicht berücksichtigt, ergibt sich (fälschlicherweise) ein Zusammenhang zwischen dem Vorhandensein von gelben Finger und Lungenkrebs. Abb. 2.5 veranschaulicht das mit Hilfe eines so genannten *Confounding-Dreiecks*.



**Abb. 2.5:** Schematische Darstellung von Confounding am Beispiel der Assoziation von gelben Fingern und Lungenkrebs; nach: Davey Smith G, Phillips AN. Confounding in epidemiological studies: why “independent” effects may not be all they seem. *British Medical Journal* 1992; 305: 757–759.

Häufige Confounder in epidemiologischen Untersuchungen sind das *Alter* und der *sozioökonomische Status*. Es gibt verschiedene Möglichkeiten, potenzielle Confounder bei der Durchführung und Analyse von Studien zu berücksichtigen.

- So kann man bereits im Studiendesign vorsehen, nur Personen in die Studie einzuschließen, die im Hinblick auf mögliche Confounder homogen sind. Man würde dann z.B. nur Personen ähnlichen Alters oder mit einem ähnlichen sozioökonomischen Status in die Studie aufnehmen (Einschränkung der Studienbevölkerung).
- In Fall-Kontroll-Studien besteht darüber hinaus die Möglichkeit, für potenzielle Confounder zu *matchen* (*Matching*). Hierzu wird jedem Fall eine Kontrolle zugeordnet, die ihm in der Ausprägung eines oder mehrerer Confounder ähnelt.
- In randomisierten kontrollierten Studien (s. Kap. 2.1.6) werden Personen nach dem Zufallsprinzip der Interventions- und Kontrollgruppe zugeordnet. Auch potentielle Confounder werden dadurch gleichmäßig zwischen beiden Gruppen verteilt, sodass *Strukturgleichheit* entsteht.
- Wenn eine epidemiologische Studie bereits durchgeführt wurde, ist es möglich, Confounding durch eine *stratifizierte Analyse* zu kontrollieren. Dazu wird die Assoziation

zwischen Exposition und Outcome getrennt für die einzelnen Ausprägungen (*Strata*; von lat. *Stratum* = Schicht) des vermeintlichen Confounders ermittelt. Confounding liegt dann vor, wenn die ermittelte Assoziation größer oder kleiner ist, als bei der unstratifizierten Analyse errechnet wurde.

- Werden mehrere Confounder vermutet, können diese durch *multivariate Verfahren* kontrolliert werden.

Die beiden letzten Möglichkeiten setzen voraus, dass potenzielle Confounder in der Studie erhoben wurden.

Variablen, die Zwischenstufen auf dem kausalen Pfad von Exposition und Outcome darstellen – so genannte *Intermediärvariablen* – sind keine Confounder. Ein Beispiel hierfür ist „geringes Geburtsgewicht“ bei Kindern von Frauen, die während der Schwangerschaft geraucht haben. Die *Exposition*, das Rauchen während der Schwangerschaft, führt hier über ein geringes Geburtsgewicht (*Intermediärvariable*) zu einer erhöhten Sterblichkeit der Säuglinge in der ersten Lebenswoche (*Outcome*).

Effektmodifikation

Die Stärke der Assoziation zwischen Exposition und Outcome kann sich je nach Ausprägung einer dritten Variablen – dem so genannten *Effektmodifikator* – unterscheiden. Dies wird als **Effektmodifikation** oder *Interaktion* bezeichnet. Tab. 2.4 zeigt die Ergebnisse einer Meta-Analyse, in der untersucht wurde, ob es einen Unterschied im Zusammenhang zwischen der *Exposition* Hepatitis-B-Virus-Infektion und dem *Outcome* Leberkrebs bei Rauchern und Nichtrauchern gibt. Die Vergleichsgruppe besteht bei allen angegebenen Relativen Risiken aus NichtraucherInnen, die nicht mit dem Hepatitis-B-Virus (HBV) infiziert, also HBV-negativ, sind.

**Tab. 2.4:** Relatives Risiko für die Entwicklung von Leberkrebs in Abhängigkeit von einer vorhe-rigen Hepatitis-B-Virus-Infektion (Exposition) und von der Frage, ob die betroffene Person Raucher ist (Effektmodifikator); Quelle der Originaldaten: Chuang SC, Lee YC, Hashibe M, Dai M, Zheng T, Boffetta P. Interaction between cigarette smoking and hepatitis B and C virus infection on the risk of liver cancer: a meta-analysis. Cancer Epidemiology, Biomarkers and Prevention 2010; 19: 1261–1268.

		Effektmodifikator (Rauchen)	
		Nein	Ja
Exposition (Hepatitis-B-Virus-Infektion)	Nein	1,0 (Referenz)	1,9
	Ja	15,8	21,6

Die Untersuchung zeigt, dass NichtraucherInnen, die mit dem Hepatitis-B-Virus (HBV) infiziert sind, im Vergleich zu HBV-negativen NichtraucherInnen ein 15,8-fach höheres Risiko haben, ein Karzinom der Leber zu entwickeln. Das Risiko von HBV-positiven RaucherInnen, ein Leberkarzinom zu bekommen, ist jedoch mit einem RR von 21,6 noch deutlich höher. Beide Faktoren (HBV-Infektion und Rauchen) wirken also zusammen und verstärken sich gegenseitig. Anders als Confounding und Bias ist eine Effekt-

modifikation keine Verzerrung. Es ist jedoch ebenso wichtig, sie zu identifizieren und in der Auswertung zu berücksichtigen, da sonst wichtige Zusammenhänge unerkannt bleiben. Effektmodifikationen können wie Confounding mittels stratifizierter oder multivariater Analyse aufgedeckt werden.

### Nur Assoziation oder auch Ursache?

Um Empfehlungen für Public-Health-Maßnahmen aussprechen zu können, reicht es nicht aus, die Assoziation zwischen Exposition und Outcome fehler- und verzerrungsfrei zu bestimmen. Zusätzlich muss ermittelt werden, ob ein Ursache-Wirkungs-Zusammenhang zwischen Exposition und Outcome besteht, die Assoziation also auch kausal ist. Einen statistischen Test auf Kausalität gibt es nicht. Daher müssen alle Hinweise für und gegen einen kausalen Zusammenhang sorgfältig beurteilt werden. *Sir Austin Bradford Hill* benannte schon 1965 verschiedene Kriterien, die es erleichtern, eine Assoziation auf ihre Kausalität hin zu beurteilen. Sie werden als **Bradford-Hill-Kriterien** bezeichnet:

- *Stärke der Beziehung*: Die Wahrscheinlichkeit einer kausalen Beziehung zwischen Exposition und Outcome nimmt mit der Stärke der Assoziation zu.
- *Konsistenz der Beziehung*: Wird eine Assoziation in mehreren unterschiedlichen Bevölkerungen (in verschiedenen Ländern, bei Personen unterschiedlichen Alters etc.) und mittels unterschiedlicher Studientypen festgestellt, ist eine kausale Beziehung wahrscheinlicher, als wenn dies nicht der Fall ist.
- *Zeitliche Sequenz*: Bei einer kausalen Beziehung muss die Ursache der Wirkung zwingend vorausgehen. Querschnittstudien sind daher weniger gut geeignet als Kohorten- und experimentelle Studien, auf eine eventuell vorhandene Kausalität zu schließen, da sie Expositionen und Assoziationen gleichzeitig erheben.
- *Spezifität des Effekts*: Dieses Kriterium fordert, dass eine Exposition (z. B. das Masernvirus) nur mit einem einzigen Outcome (hier: einer Masernerkrankung) assoziiert ist. Für die Untersuchung von Expositionen wie Rauchen oder Alkoholkonsum, die das Risiko für viele verschiedene Outcomes erhöhen, ist es weniger hilfreich.
- *Dosis-Wirkungs-Beziehung*: Steigt mit zunehmender „Menge“ der Exposition (z. B.: 1–10 Zigaretten/Tag, 11–20 Zigaretten/Tag, 21–30 Zigaretten/Tag) das Risiko, dass der Outcome eintritt, ist eine Kausalität wahrscheinlicher als ohne Dosis-Wirkungs-Beziehung.
- *Biologische Plausibilität und Kohärenz*: Eine Ursache-Wirkungs-Beziehung zwischen Exposition und Outcome sollte biologisch plausibel sein und nicht im Widerspruch zu den Ergebnissen anderer Fachgebiete stehen.
- *Experimentelle Evidenz*: Kann man mittels einer randomisierten kontrollierten Studie zeigen, dass das Risiko für einen Outcome sinkt, sobald die Exposition beseitigt ist, liegt sehr wahrscheinlich eine Ursache-Wirkungs-Beziehung vor.

Die Bradford-Hill-Kriterien können die Beurteilung einer Ursache-Wirkungs-Beziehung zwischen Exposition und Outcome unterstützen, sie sollten jedoch nicht als Checkliste missverstanden werden. Nur selten wird eine Ursache-Wirkungs-Beziehung so deutlich



wie bei der Assoziation zwischen Rauchen und Lungenkrebs. Hier wurden mit Ausnahme der *Spezifität des Effekts* alle oben genannten Kriterien erfüllt.

### Internet-Ressourcen

Auf unserer Lehrbuch-Homepage ([www.public-health-kompakt.de](http://www.public-health-kompakt.de)) finden Sie Hinweise auf Quellen und weiterführende Literatur, Vorlesungen sowie Links zu den erwähnten Studien und Institutionen, u. a. zu Todesursachenstatistiken aus Deutschland und der Schweiz.

## 2.2 Demografie

*Marcel Zwahlen, Matthias Egger, Johannes Siegrist*

Die Frage „Wie viele sind wir?“ bewegt Regierungen bereits seit dem Altertum. Sie bildet die Grundlage der *Demografie* [von *démos* (gr.): Volk und *grafé* (gr.): Schrift, Beschreibung], die sich mit verschiedenen Merkmalen von Bevölkerungen beschäftigt. Dabei interessieren neben der Gesamtgröße der Bevölkerung, ihrer altersmäßigen Zusammensetzung und ihrer geografischen Verteilung auch die sozialen und Umweltfaktoren, die hier für Veränderungen verantwortlich sind. Die Daten zur fortlaufenden Beschreibung der Bevölkerung stammen mehrheitlich aus staatlichen Quellen, v. a. aus Volkszählungen, dem Geburten- und Sterberegister sowie repräsentativen Stichproben-Erhebungen.

In diesem Abschnitt beschäftigen wir uns mit den Kennziffern, die Demografinnen zur Beschreibung einer Bevölkerung verwenden, z. B. dem Geburtenüberschuss, dem Wanderungssaldo, verschiedenen Sterberaten, der Lebenserwartung und potentiell verlorenen Lebensjahren. Abschließend betrachten wir häufig verwendete grafische Darstellungen, z. B. zur Altersstruktur einer Bevölkerung und erläutern zeitliche Trends in West- und Ostdeutschland sowie in der Schweiz.

Schweizerische Lernziele: CPH 17–20

### 2.2.1 Die Bevölkerung

Das Lukasevangelium berichtet über eine Anordnung des römischen Kaisers Augustus, nach der sich alle Bewohner des Reiches für eine Volkszählung in ihre Herkunftsorte zu begeben hatten. Maria und Josef reisten daraufhin nach Bethlehem, wo Jesus geboren wurde. Die Registrierung der Bevölkerung gab den Verantwortlichen in Rom einen Überblick über die Anzahl ihrer Steuerbürger. Die einfachste Information über eine Bevölkerung bezieht sich also auf die Zahl der Personen, die sich in einer geografisch definierten Region an einem bestimmten Datum befinden. Doch wenn Sie beispielsweise am 15. Juli die Anzahl an Personen zählen, die sich auf Mallorca befinden, erhalten Sie möglicherweise nicht die Zahl, die in *Wikipedia* unter „Bevölkerung von Mallorca“ aufgeführt wird (862.397 Einwohner für das Jahr 2009). Denn der Monat Juli ist Ferienzeit, und Sie werden Personen zählen, die nicht auf Mallorca wohnen, sondern